0001

IN THE UNITED STATES DISTRICT COURT FOR THE
NORTHERN DISTRICT OF OKLAHOMA

W. A. DREW EDMONDSON, in his capacity as ATTORNEY GENERAL OF THE STATE OF OKLAHOMA and OKLAHOMA SECRETARY OF THE ENVIRONMENT C. MILES TOLBERT, in his capacity as the TRUSTEE FOR NATURAL RESOURCES FOR THE STATE OF OKLAHOMA,	)
Plaintiff,	)
vs.	)4:05-CV-00329-TCK-SAJ
TYSON FOODS, INC., et al,	)
Defendants.	)

- -

VOLUME I OF THE VIDEOTAPED
DEPOSITION OF CHARLES COWAN, PhD, produced as a witness on behalf of the Plaintiff in the above styled and numbered cause, taken on the 17th day of February, 2009, in the City of Tulsa, County of Tulsa, State of Oklahoma, before me, Lisa A. Steinmeyer, a Certified Shorthand Reporter, duly certified under and by virtue of the laws of the State of Oklahoma.

0002

A P P E A R A N C E S

FOR THE PLAINTIFFS:	Mr. David Page Attorney at Law 502 West 6th Street Tulsa, OK 74119
FOR TYSON FOODS:	Mr. Gordon Todd Attorney at Law 1501 K Street N.W. Washington, D.C. 20005
FOR CARGILL:	Ms. Theresa Hill Attorney at Law 100 West 5th Street Suite 400 Tulsa, OK 74103 -and- Ms. Melissa Collins Attorney at Law 1700 Lincoln Street Suite 3200 Denver, CO 80203
FOR SIMMONS FOODS:	Mr. Bruce Freeman Attorney at Law One Williams Center Suite 4000

Exhibit B

Tulsa, OK 74172

FOR GEORGE'S: Ms. K. C. Tucker
Attorney at Law
221 North College
Fayetteville, AR 72701

FOR CAL-MAINE: Mr. Robert Sanders
Attorney at Law
2000 AmSouth Plaza
P. O. Box 23059
Jackson, MS 39225
(Via phone)

ALSO PRESENT: Roger Olsen, PhD

I N D E X

(Whereupon, the deposition began at 9:09 a.m.)

VIDEOGRAPHER: We are now on the Record for the deposition of Dr. Charles Cowan. Today is February 17th, 2009. The time is 9:09 a.m. Would counsel please identify themselves for the Record? 09:09AM

MR. PAGE: David Page for the State of Oklahoma, and with me here today is Dr. Olsen, an expert for the State of Oklahoma.

MR. TODD: Gordon Todd for the Tyson Food Companies. 09:10AM

MS. COLLINS: Melissa Collins for the Cargill defendants.

MS. HILL: Theresa Hill for the Cargill defendants. 09:10AM

MR. FREEMAN: Bruce Freeman for Simmons.

MR. TUCKER: K. C. Tucker for the George's defendants.

VIDEOGRAPHER: And on the phone?

MR. SANDERS: Bob Sanders for the Cal-Maine 09:10AM

defendants. I think I'm the only one.
VIDEOGRAPHER: Thank you. The witness may be sworn in.
CHARLES COWAN, PhD
having first been duly sworn to testify the truth, the whole truth and nothing but the truth, testified as follows:
DIRECT EXAMINATION
BY MR. PAGE:
Q Would you state your full name for the Record, please? 09:10AM
A Charles Douglas Cowan.
Q And what is your address?
A Work or home?
Q Both, please. 09:10AM
A Okay. Home address is 5218 Sagail Place. Sagail is S-A-G-A-I-L Place, San Antonio, Texas 78249. My office address is 4939 De Zavala Road. D-E one word. Separate word is Zavala, Z-A-V-A-L-A. And that's also in San Antonio, Texas 78249. 09:11AM
Q Have you ever had your deposition taken before, Dr. Cowan?
A Yes, sir.
Q And when was that?
A Well, it's actually 30 or 40 times. 09:11AM
Q Okay. When was the most recent time?
A Two weeks ago.
Q In what matter was that?
A It was -- sorry. Moregate versus Mailboxes, Etc. It's in southern California. 09:11AM
Q Okay, and can you tell me what the general nature of that litigation is involving?
A Sure. When -- several years ago UPS bought Mailboxes, Etc. Several of the franchisees for Mailboxes, Etc., felt that the purchase wasn't in their best interest, that they weren't being adequately compensated or represented by the new combined entity, and so they are suing for lost profits and lost business opportunities. 09:12AM
Q And that case does not involve environmental matters; correct?
A No, it does not.
Q Have you ever been deposed in a case that involves environmental matters?
A Several times. 09:12AM
Q Okay. Could you identify those for us, please?
A Sure.
Q And when you do that, if you could just tell us the type of environmental issues involved briefly, that would be help -- be helpful. 09:12AM
A Sure. Most of the cases have involved groundwater or airborne contamination around a plant or a -- some other type of facility that had some type of discharge. In those cases, the contaminant was typically something like fertilizer that had leached into groundwater, had been spreading over time, and the claims were that the contamination diminished the value of properties that were in the path of the groundwater. 09:13AM

Q And was your role an economic analysis or an environmental analysis in those cases?
A Economic.
Q Have you had any cases where you've actually done an environmental analysis as an expert? 09:13AM
A No.
Q So this is your first case where you've done an environmental statistical analysis as an expert?
A I'm not sure how to understand your question.
Q Well, I just -- you testified that the four or five cases that you've been deposed involving groundwater and airborne contamination, you were doing an economic analysis for the litigants in that case; correct? 09:14AM
A Yes. 09:14AM
Q In this particular case, are you doing an economic analysis?
A No.
Q Okay. Aren't you evaluating statistically the environmental data that's associated with the claims 09:14AM

0008

in this case?
A No.
Q What are you doing in this case?
A I'm evaluating the quality of the statistical analysis that was done by Dr. Olsen. I'm not doing 09:14AM a separate statistical analysis.
Q Okay.
A And then to answer the first question you asked, in each of those cases, I had to determine what was the environmental impact, what was the 09:14AM spread of the contaminants. Plus, you didn't allow me to finish my description. So in those cases, you couldn't do the economic analysis absent any knowledge of what the environmental contamination was. 09:15AM
Q But in those cases, and I'm just trying to broad brush it. If not, we'll go individually. In those cases, were you personally evaluating the sources of contamination and the scope and extent of the contamination? 09:15AM
A No.
Q So you relied on the statements of other experts and then did your evaluation; correct?
A I did.
Q Okay. So what I'm trying to hone in on here, 09:15AM

0009

Dr. Cowan, is whether or not this case is the first time that you've actually evaluated the environmental data from a statistical perspective?
A And I just answered that question and said no, it's not. In each of the other cases I had to 09:15AM evaluate the environmental data that I was given and work with hydrologists and experts like that to be able to determine what they were telling me and what their analysis was before I could conduct my analysis. 09:15AM
Q In these previous cases, did you actually critically review the environmental data; that is, did you look at the statistical analysis provided by the experts that were identifying sources in those cases and do a critical review in those cases? 09:16AM
A I did because, otherwise, I couldn't know how

valid or reliable my economic analysis was.
Q Okay. Would you tell me about the first case in the most recent past that involved either -- you said there was four or five, so let me go through those. Let's go from the most recent and go backwards. Okay? 09:16AM
A Okay.
Q So what would be the most recent case you've -- involving environmental contamination you've worked on? 09:16AM
A There was a case involving Conoco in Pensacola, Florida, where it was Conoco, Agrico and a third company that had gone out of business, so it was primarily Conoco and Agrico. They jointly operated a site which produced fertilizer, among other things, and they -- over time rainwater or rain had caused fertilizer to go into the groundwater and then had spread through the area where -- in Pensacola down into a large bayou, which fronted onto the ocean, but the bayou was important because of all the properties that ringed the bayou having unique values relative to the rest of the city. 09:16AM 09:17AM
Q Okay. In that case did you do a critical or were any of your opinions -- let me strike that. In that case did you offer any opinions as to the source of the contamination? 09:17AM
A Well, that source was a given because of the nature -- 09:18AM
Q So the answer is no?
A -- of the lawsuit. No.
Q Okay, and in that case did you offer any opinions concerning the fate and transport of the contamination that was involved? 09:18AM
A I did.
Q And what was your opinion involving that case?
A Well, there were actually two analyses done, one for the plaintiffs and one for the defendants.
Q And you were working for who? 09:18AM
A The defendants.
Q Okay, and what was your analysis with regard to fate and transport in that case?
A Well, the problem was that the two analyses were so incredibly different from one another, that I had to determine what was a reasonable analysis and what was a reasonable analysis on their point that could then be used to determine the likelihood of diminution of value in properties, and so I was contrasting and working with the two opinions or the two reports to come to some midpoint. 09:18AM 09:19AM
Q Okay. So you tried to determine what the central tendencies of each of the opinions is so you could come up with a mean or a midpoint between those two? 09:19AM
A A little broader than that because I needed to know how reliable. It wasn't so much the central tendencies because both reports agreed on that. It was where the edges were.
Q Okay. Did you actually critically review the analysis of fate and transport of the fertilizer in 09:19AM

the groundwater or were you simply given that as the two different sides, opinions and try to determine what the central tendency -- or excuse me, what the midpoint was between the two? 09:19AM

A I was given the reports and I analyzed those.

Q Okay. So you took the data. You didn't actually express an opinion on whether or not fertilizer actually did move in a certain direction in the groundwater from the plant in question, did you? 09:20AM

A Not in that case.

Q Okay. In front of you -- could you identify what the exhibit in front of you is marked as Cowan Exhibit No. 1 right here? 09:20AM

A That's my rebuttal report.

MR. TODD: Take a minute to just flip through it.

Q Yeah. You might want to take a moment just to make sure because I may characterize something, but I want to make sure that you agree with my characterization. 09:20AM

A Yes, sir.

Q And while you're going through there, what I want you to do is, if you would for me, identify in 09:20AM

0013

the report any reference you have, maybe in your experience or CV, that discusses the case that you just mentioned.

A Okay. I've read through the report. It is, as nearly as I can tell, my complete report. If you 09:21AM go to Page 71, which is the second to the last page in the report, I list jointly three cases that were property value diminution cases and the last one listed is Bernice Samples versus Conoco, Agrico and Escambia Treating. That was the case we were just 09:21AM discussing.

Q Excuse me a second. It turns out the copy I had in front of me didn't have Pages 71 and 72.

MR. TODD: David, is this an additional copy? 09:22AM

MR. PAGE: Yes, that is. Now this one doesn't have 71 or 72.

Q Could you then direct my attention on 71?

A 71, the third to the last paragraph, toxic tort, the last two full lines -- well, the last 09:22AM three full lines, Bernice Samples versus Conoco, Agrico and Escambia Treating, is the case we were just discussing.

Q So in that case you were offering opinions on diminution in value; correct? 09:22AM

0014

A Among other things, yes.

Q Well, did you actually testify in court in that case?

A No. Well, there was a deposition. It didn't go to trial. 09:22AM

Q Okay. Is it still pending?

A No. It settled.

Q Okay. Now, the next most recent case, again, involving environmental matters, if you could, identify that for us, please, sir. 09:23AM

A There was a case before that also in Florida that was also a toxic tort case. It was actually

quite similar. It also involved Agrico, but it was in Lakeland, Florida and, again, it had to do with fertilizer and contamination of groundwater. 09:23AM

Q And what were your opinions in that case?

A Similar, in that I was looking for diminution in value.

Q Okay. So your primary focus was to evaluate the diminution in value of the property in both of these cases, was it not? 09:23AM

A It was, although I'd like to correct something I said a minute ago. I'd not thought about this, but this will come up in the third case, too. In terms of sources, I was also -- as part of the 09:23AM

0015

analysis that I conducted, I had to look at sources because in Pensacola, there was a large naval base which was also a source of groundwater contamination.

Q Okay, but did you -- were you the expert that was principally involved with identifying what or which were the sources of contamination in those cases? 09:24AM

A Well, I was one of them in terms of -- my interest and my involvement had to do with the diminution in value as opposed to the environmental --

Q Right, but if I got a copy of those reports in that case, would it identify an analysis by you of which were the primary sources of the contamination and your basis for that? 09:24AM

A If you mean from an environmental perspective --

Q Yes.

A -- no. From an economic perspective, yes. 09:24AM

Q Okay. So from an environmental perspective, you didn't identify sources in any of these cases; is that correct?

A In the two cases we've discussed so far.

Q Okay, and can you identify this Lakeland, 09:24AM

0016

Florida case discussion in your Exhibit 1 to this deposition?

A It's also in the same paragraph on Page 71.

Q So you refer to these as toxic tort in your CV; correct? 09:25AM

A Yes, sir.

Q Okay. The next case, sir?

A It goes --

Q My count it's the third case.

A Mine, too. 09:25AM

Q Good.

A Excuse me. I'm thinking about timing so I can get this chronologically.

Q If you don't get it perfect, that's okay.

A Okay. Thank you. Because there are two cases at about the same time but they were quite different from one another. The -- excuse me. They're in St. Petersburg, Florida, Pinellas County. There was a phosphorus plant owned by a company called Stouffer, spelled like the food company. This was a class 09:25AM action against Stouffer because Stouffer had purchased the phosphorus company, and under Florida state law they had purchased it for the purpose of

cleaning it up, and then they were going to resell it, but their primary mission in life was to remediate environmental properties. 09:26AM

During the cleanup of the phosphorus, the phosphorus exploded and there was a huge cloud of phosphorus in the air. It -- there was airborne contamination, and the question was both -- well, primarily diminution in value for the properties that were around this phosphorus plant. 09:26AM

Q And was that the primary focus of your opinion in those two cases, the diminution in value of the property? 09:26AM

A Okay, but we're up to three.

Q Oh, I'm sorry. You said there were two similar. So we're only talking about one now.

A Oh. Just the phosphorus case, yes.

Q Okay. So St. Petersburg, Florida was the third case? 09:27AM

A Yes, sir.

Q Was a phosphorus plant where the purchaser was to remediate the facility; correct?

A Yes. 09:27AM

Q And there was an explosion?

A Right.

Q In that case was your primary focus of your opinion the diminution of value of the properties surrounding the plant? 09:27AM

A Yes, sir.

Q Okay. Did you do any evaluation as to the scope and extent, that is, were you primarily responsible for the evaluation and scope and extent of the contamination that was involved in that case? 09:27AM

A No.

Q What's -- is that one -- is that particular case identified in your CV, sir?

A That's the third one listed under the heading toxic tort. 09:27AM

Q Thank you, sir. Okay. Can we go to No. 4, please?

A Sure. In Scottsdale, Arizona, there was a plant -- this was a long time ago, so I don't think this is a secret anymore. Motorola has a plant where it produces circuit boards, and for the circuit boards -- once the circuit boards are etched, they're cleaned with a chemical solution, and the chemical solution ran into the groundwater. The plant had been in operation for 40 years. 09:28AM

Q Do you know what chemical solution was involved?

A I don't remember off the top of my head.

Q That was the principal contaminant?

A Yes. 09:28AM

Q You don't recall what the contaminant was?

A Well, we're talking about fifteen years ago.

Q Okay.

A So if I were allowed to go back and look at my records, I would, but I don't. 09:28AM

Q I'm just checking --

A Okay.

Q -- what you understood today. So you -- the

issue was the groundwater contamination of some cleaning elements for the circuit boards at the Motorola plant? 09:28AM

A Yes.

Q Okay.

A And then the EPA one day decided that that chemical was a -- the chemical in the cleaning solution was a carcinogen, and so there were the beginnings of a class action suit being filed against Motorola for contaminating the groundwater, and I was asked to determine the likelihood of -- the likelihood and number of people who were exposed from a medical perspective to this carcinogen and what would be the likely outcomes. 09:28AM 09:29AM

Q Okay. So you -- would you characterize your analysis as epidemiological in that particular case or how -- what would you characterize that? 09:29AM

0020

A It was a combination of epidemiology and demography.

Q Demography, okay. Did any of your work in that case or, excuse me, your opinions in that case involve determination of scope and extent of these cleaning solvents in the environment? 09:29AM

A I don't know what you mean by scope, but certainly the extent.

Q Okay, but you didn't do that yourself; you relied on other experts to tell you how far the expanse was of the contaminants in the groundwater; is that not correct? 09:30AM

A Well, I worked with them, yes, but I relied -- I relied on the work that they did. I worked with them as they were beginning to get into this. 09:30AM

Q But you weren't the one that modeled, for example, the cleaning solvents in the groundwater; correct; you didn't do that analysis?

A Well, I'm having trouble responding to your question because if you're talking about modeling of the cleaning solvents in the water, no. If you're talking about the extent of the dilution and how far out it spread, yes. 09:30AM

Q You did the calculations on the dilution?

A I worked with the hydrologists on it. 09:30AM

0021

Q Okay. Were you -- did you give an opinion on the solution, or did the hydrologists provide the opinion on the dilution of these contaminants in the groundwater?

A I gave a slightly different opinion in terms of the impact of the solution after I relied on it from the -- 09:30AM

Q Right?

A Okay.

Q Okay. And that analysis provided or is that case discussed in your CV that's in Exhibit No. 1? 09:31AM

A No.

Q Why not?

A It never got far enough that the -- that it was filed. There was just initial discussions about it. So I was hired to do the epidemiological work. 09:31AM

Q So that was your primary focus was epidemiology in that case?

A Well, that and the demography. You couldn't

-- the two different -- it's two different bags of tools. 09:31AM

Q When you say demography, you're talking about the characteristics of the populations of individuals or people in the area?

A Yes, sir. 09:31AM

Q Thank you. Now, did you give your deposition in that fourth case from Scottsdale, Arizona?

A No, I did not.

Q Okay. In these first four cases we've discussed, did you provide a written report? 09:31AM

A Yes.

Q Do you still have those written reports?

A I'm not sure about the Pensacola case, and the other three, no.

Q Okay. Would you have any objections to checking and providing those to your counsel so you could provide me copies of any reports you still have available? 09:32AM

A I'd be happy to.

MR. PAGE: I'd like to make that request. 09:32AM

MR. TODD: Sure. I'd just ask that you put it in writing after the deposition.

MR. PAGE: You bet.

MR. TODD: We'll be happy to look into that. 09:32AM

MR. PAGE: You bet.

Q On the other three cases, on the Conoco case I think you mentioned it was fertilizer. What were the chemicals of concern in the first case we talked about, the one that's just recently? 09:32AM

A The Pensacola case, that's the most recent case.

Q Yes, sir. What did I say? Did I say Agrico? Excuse me.

A No. That's okay.

Q Conoco and the Pensacola, yes, sir.

A Right. Okay. Well, there was -- this is one of the reasons why there was some source confusion in this case. The primary concern about the fertilizer was ammonia. However, the problem in the 09:33AM groundwater contamination that was discovered after you got up to the bayou was uranium, which is -- as far as we could tell wasn't part of the production process for Conoco or Agrico.

Q So when you did your evaluation of diminution of value, which chemical were you considering? 09:33AM

A Well, as an economist, you wouldn't consider one specific chemical. You would consider their cumulative effect.

Q Okay.

A And what impact they had on the values of the properties.

Q So you were acting as an economist in that case?

A Yes. 09:34AM

Q Okay, and on the Agrico-Lakeland case, that was I think the second one we talked about?

A Yes, sir.

Q What were the chemicals of concern in that

case?
A Same issue because it's fertilizer. So, again, the primary one I remember is ammonia, but there was no uranium involved in that one.
Q Okay, and what about the St. Petersburg, Florida plant; what were the chemicals of concern involved in that case? 09:34AM
A Well, since it was a phosphorus plant, phosphorus.
Q It was phosphorus, okay. And was there any residual phosphorus in the environment that you evaluated or was it simply the effects of the initial explosion that you were concerned with in that case? 09:34AM
A I don't know how to answer your question because are you talking about residual phosphorus as phosphorus or are you talking about residual phosphorus after it's combined with something else? 09:34AM
Q Yeah, after it's combined, the results of the combustion.
A Okay. That's good because if it hadn't 09:35AM

combined, it would still explode.
Q Yeah, well, it wouldn't be in the environment naturally, would it be, phosphorus?
A No, because if it --
Q If it's exposed to air, it immediately combusts; correct? 09:35AM
A Yes.
Q Okay. So what were the chemicals of concern after the explosion in the St. Petersburg, Florida plant? 09:35AM
A I don't recall.
Q Okay, and Scottsdale, you just remember it was a cleaning agent; you don't recall what it was?
A No. In both of these cases we're talking fifteen years ago, so -- 09:35AM
Q And you also -- okay, and there was a fifth case you said that involved some environmental contamination involvement.
A This was a case involving a dry cleaner and the remediation of or the -- how -- it's not a single shop. It's a large chain of dry cleaners and how they dealt with the requirements to take care of the discharge from dry cleaning. 09:35AM
Q Okay, and do you remember the location where this case occurred? 09:36AM

A Florida.
Q Florida, okay. And what was your role in that case, sir?
A I was supposed to determine whether or not the cleaner had been deceptive in the way that they worked with both the State and with their consumers. So it was a deceptive sales practices case in terms of how they worked with the State and the consumer in the way they dealt with the contaminants that would result from dry cleaning. 09:36AM
Q Okay. Did your work in that case involve an evaluation of the scope and extent of contamination?
A No.
Q Do you recall where the contamination was in that case? 09:36AM

A Well, what I said was --
Q It was more a record keeping kind of a case; is that what it was?
A It was more of a record keeping case because it was every dry cleaner for this large corporation, but we're talking about hundreds of locations. 09:37AM
Q So your evaluation was more of a records analysis to see if they properly reported their disposal or management of their cleaning fluids?
A No. It was actually how they dealt with the State in terms of the reporting to the State about the costs for remediation, what they had done to adhere to state law and then how they dealt with that in their pricing for consumers. 09:37AM
Q But was it mostly evaluation of their records -- of what they told the State through their records?
A Well, told the State and then told consumers also. So there was two different sides to this.
Q But just to make sure, it did not involve an evaluation of the contamination at these particular dry cleaning locations? 09:37AM
A No.
Q Any other cases involving environmental matters? 09:37AM
A Not that I recall.
Q Okay, and the fifth case we just talked about, is that reported in your CV, sir?
A I believe it is.
Q Can you show me where? 09:38AM
A Yes, sir. Page 70.
Q Under deceptive sales practices?
A Yes, sir, the second one, Watkins versus Dry Cleaners International.
Q Looking through your CV, I just don't sense that there's a lot of experience you have working with contaminants in the environment. Is that a fair characterization? 09:38AM

MS. HILL: Object to form.

A Of course, that wasn't why I was hired, so --
Q Can you answer the question yes or no? 09:38AM
A No, there's not a lot of experience dealing with the determination of environmental contaminants and their sources.
Q Other than the description of these five cases that you just provided us, can you tell me if you have any other experience, whether it's involved in a case or not, not necessarily litigation -- I'm trying to look at experience beyond litigation -- where you've done evaluation of datasets that involve geochemical or environmental data? 09:39AM
A If you -- are you using the -- I understand the geochemical. Are you using environmental in the narrow sense of relating to how it affects the earth as opposed to environmental in terms of sociological concerns? 09:39AM
Q Yes, sir.
A Okay. Then, no, I have not had any other involvement.
Q Okay. So this would be your first case where 09:39AM

you evaluated such a dataset as in this case?
A Well, keep in mind, I didn't evaluate the dataset. I evaluated Dr. Olsen's work.
Q Well, you did, though, did you not, comment on whether or not Dr. Olsen's dataset was reproducible; 09:39AM
correct?
A Yes, I did.
Q Okay. So I guess let me restate the question this way: Is this the first time -- I hope there's
no underlying -- I'm trying to make this as simple 09:40AM
as possible. Is this the first dataset that you've evaluated that deals with environmental data defining environmental data the way you just did?
A Okay. Well, I want to be able to distinguish
between evaluating the data itself, which I didn't 09:40AM
look at, versus evaluating Dr. Olsen's data because he constructed his datasets from that original dataset.
Q Okay. Let me ask you this question then.
Maybe this is a better question. Is this the first 09:40AM
case where you've done a review of statistical analysis of how another expert did statistical analysis on an environmental dataset?
A Yes, it is.
Q Thank you. I knew if I got enough tries, I 09:40AM

could ask a good question --
A Thank you, sir.
Q -- that got to the point. If you bear with me here today --
A And I appreciate it. 09:41AM
Q Thank you. Have you ever -- I assume this is the case. Have you ever done any microbial source tracking work?
A Well, I'm not exactly sure how to answer that
question only because I'm not sure how you 09:41AM
characterize the work I did. So if I could describe a case that involved microbial source tracking, I worked on a case involving barges on the Mississippi River.
Q Yes, sir. 09:41AM
A And the question was whether or not the materials used to coat the interior of the barges' holds were adequate to keep bacteria from eating into the hulls of the boats. So what happened was
that there were a series of experts pulled together, 09:42AM
some who were microbiologists, some who were geochemists, some who were engineers, and each person was involved in some aspect of collecting and organizing data on what the coatings were in the
barges on the Mississippi, how intact were they, the 09:42AM

conditions within the holds and finally the extent of pitting that had been in the barges, pitting being sort of eating away of the interior of the hull.
My job was to coordinate the job of everybody 09:43AM
else and then analyze the data they collected. So I helped with the front end in terms of thinking about how one goes about collecting the data and what was a representative sample. I worked with --
Q So that was -- so in that case, was the issue 09:43AM
the source of bacteria that was intruding into

containers on a ship?
A Yes, sir.
Q Okay.
A Okay. 09:43AM
Q So that was the -- the bacteria you're looking at to see whether or not there was bacteria on a ship getting into containers that were being transported by that ship; correct?
A Well, not necessarily because the question -- 09:43AM
part of the question was what had the barge owners done that would encourage the growth of bacteria, and so there's -- so there are a lot of different sources of bacteria, and the question was whether or
not they had done a sufficient amount to protect the 09:44AM

0032

interior of the boat over and above the covering as opposed to the invasion, and then the question -- the secondary question was, did it matter what part of the Mississippi, did it matter whether it was
saltwater or not, did it matter what the boats were 09:44AM
hauling, did it matter what the configuration of the boats were. So there were a lot of other factors that went into --
Q Is that the only experience you've had with
bacteria source tracking? 09:44AM
A Oh, no.
Q Have you ever worked in a case where there's bacteria source tracking in the ambient environment, such as the issues involved in this case?
A Actually I'm working on a project that's not a 09:44AM
case, but I'm working on two projects right now that involve the spread of different types of diseases. One is in Lima, Peru, where I'm working to study the spread of multidrug resistant tuberculosis
throughout the population in Lima that would be 09:45AM
sourced at prisons, and then the prison structure in Lima is quite a bit different than it is here so that you have --
Q So you're looking at whether or not there is
contaminated food and contaminated -- 09:45AM

0033

A No.
Q Well, so is that -- is it concern about bacteria in a prison; is that what the concern is?
A No. The concern is the bacteria and how it
spreads through the population outside of the 09:45AM
prison.
Q But it's people -- that people spread --
A Could you not interrupt me, please?
Q Excuse me.
A And I apologize. I don't mean to be harsh, 09:45AM
but it's just difficult for me to get my answer out.
Q That's fair enough, and I'll try not to do that.
A Thank you. Yeah, the problem is families and
the family structures and then the extended family 09:45AM
structures and then how they all interrelate so that you've got multiple pathways by which tuberculosis and other related diseases can be spread.
Q Okay.
A The other work I'm doing is for the CDC and 09:46AM
for the Bill Gates Foundation in Africa, where I've designed a research study to look at the spread of

AIDS from mother to newborn and how interventions, different interventions can effectively stop that spread from mother to newborn depending on the types 09:46AM

of drugs that are used, the care that the mother gets before the birth, the -- whether or not there is -- the mother breast feeds the baby, all the different sources of or the transmittal channels where a newborn can get AIDS from its mother, and in 09:47AM that case, I'm working with a team of pediatricians, oncologists and a variety of other doctors, but I was brought on board because they needed a statistician to coordinate the project.

Q And sometimes I interrupt, Dr. Cowan, because 09:47AM I'm thinking maybe we didn't communicate initially.

A Yes, sir.

Q I think my original question was, have you done any studies in the ambient environment? Do you understand what an ambient environment means? 09:47AM

A Could you define it for me?

Q Well, that would be outside, for example, in the fields and forests of the IRW, the Illinois River watershed.

MS. HILL: Object to the form. 09:47AM

Q That's what I mean by ambient environment.

A Well, I'm sorry. I have trouble distinguishing that between being in a city or a rural environment where -- I mean, I'm dealing with an entire country, like Zambia, where some people 09:47AM

live in the city, some people live outside, but I would consider everybody to be in an ambient environment if they're giving birth.

Q But those issues you are dealing with there, both in Africa and in Peru, isn't the focus 09:48AM person-to-person spreading of the disease?

A Well, it may or may not be depending on, first of all, the disease because tuberculosis --

Q Well, yes or no?

A Okay.

Q Is the answer then no?

A Well, I was trying to give you an answer that indicated that there is no yes or no.

Q Okay. Were those two studies primarily epidemiological studies; would you characterize them 09:48AM as that?

A I'm going to fall back to the answer I gave before on the other studies. It's a combination of epidemiology and demography.

Q Okay. Did you read Dr. Harwood's report in 09:48AM this case?

A Yes.

Q Okay. Would you -- what I'm trying to understand is if you ever reviewed any source tracking evaluation such as Dr. Harwood did in this 09:48AM

case in her report.

MS. HILL: Object to the form.

A Well, I have trouble distinguishing between what Dr. Harwood did in terms of her research and what I do in my research. I mean, if you're trying 09:49AM to make it very specific to looking at a field or a set of fields as opposed to just a general

environment, if you are talking about environment -- environmental spread, then I don't see a distinction.

Q You don't? Well, let me ask you this question: Do you not see a distinction between the spread of disease from, for example, human or animal manure being spread on fields as opposed to the type of studies you're doing in Africa and Peru currently? 09:49AM

MR. TODD: Object to form.

A Well, mathematically, no.

Q Okay, but the method -- the means of transport of the microbes are substantially different; correct? 09:49AM

A But I -- that's not my responsibility in terms of the research.

Q I understand that. That's what I'm trying to understand. 09:50AM

MS. HILL: David, would you let him finish, please?

MR. PAGE: Thank you.

MS. HILL: You're stepping all over each other. 09:50AM

A And I guess we should both apologize to Lisa. What I'm saying is that my contribution here in this case is similar to my contributions in all the research studies I've designed, which is I help evaluate whatever the pathway is, but I do it through mathematical modeling. 09:50AM

Q Okay. So you don't understand the mechanisms of bacterial source transport in the environment, do you, sir?

A Well, once again, it sort of depends on what it is we're talking about. At some point to be able to talk about the transport of the tuberculosis, I have to understand what the pathways are there and how one person can contaminate another because there's multiple pathways. 09:50AM

Q Do you consider yourself a microbiologist? 09:51AM

A No.

Q Do you consider yourself an expert in bacteria?

A No. 09:51AM

Q Have you ever designed any field sampling work to collect bacteria?

A No.

Q Have you ever designed any field sampling work to collect bacteria from manure samples? 09:51AM

A No.

Q What about land-applied fields where manure has been spread?

A No.

Q What about surface waters? 09:51AM

MS. COLLINS: Object to form.

A No.

Q Groundwater?

A Well, working with the hydrologists and describing what was needed for a representative sample, yes. 09:52AM

Q But not actually written protocols for how to sample groundwater?

A No.
MS. COLLINS: Are you specifically talking about bacteria still? 09:52AM
MR. PAGE: Yes, ma'am. All those questions were related to bacteria.
A No.
Q Would you turn to Page 40 of your report, 09:52AM

please, Footnote 16 that's on Exhibit 1. Could you read Footnote 16 for the Record, please?
A Dr. Olsen throughout his report confuses the terms parameter and variable. In this sentence he used one to explain the other. From context, it 09:52AM seems Dr. Olsen means variable when he says parameter. A parameter is the single value which describes characteristics of a population, like an arithmetic mean or a variance. A variable is the theoretical construct used to denote a value that 09:53AM can change according to the sample being observed. These are not interchangeable terms.
Q What is your concern here in Footnote 16?
A Well, the -- if you would give me one second so I can go back up to the Paragraph 87. Okay. In 09:53AM this sentence that I'm quoting from Dr. Olsen's report, he says that he is calculating a PC score using the PC coefficient multiplied by the standardized parameter concentration. This is performed for all parameters, parenthesis, 09:54AM variables, in a particular PCA run. So he uses both terms simultaneously to describe the activity that he's doing, but parameters and variables mean two completely different things.
Q So how was Dr. Olsen using them 09:54AM

inconsistently?
A Well, because you would either perform this calculation on one or the other.
Q And what would be the difference?
A Well, if you're multiplying parameters, you're 09:54AM multiplying -- using parameters the way Dr. Olsen was using parameters, you're multiplying a single number. If you're multiplying variables, you're multiplying all of the observations within one specific variable. So you could be multiplying -- 09:54AM you could be doing 597 multiplications instead of a single multiplication.
Q Aren't we talking about the individual chemicals observations when we talk about a parameter in this report, Dr. Olsen's report? 09:55AM
A No. A parameter is -- in Dr. Olsen's report is what it is that he's trying to estimate. A variable is a, what you just said, a particular chemical.
Q So it's your contention that when Dr. Olsen 09:55AM used the term parameter, he was using it as a statistical term and not as a term as environmental scientists typically use that term?
MR. TODD: Object to form.
A In the context in which he was using it, yes. 09:55AM

Q And how do you know what was in his mind?
A Well, I don't know what was in Dr. Olsen's mind. What I'm indicating is that relative to the

standard usage of the term, you can't have both at the same time. 09:55AM

Q But I think you previously testified this is the first environmental dataset you've evaluated from a statistical perspective; correct?

A Yes.

Q Are you familiar what the USGS is? 09:56AM

A Yes.

Q What is the USGS?

A U. S. Geological Survey.

Q And let me show you what's been marked as Exhibit No. 2 to your deposition. Can you identify 09:56AM that document for the Record, sir?

A It is -- seems to be a web page. At least the cover is a web page from waterdata.gs.gov that describes what you have helpfully highlighted as parameter help. 09:56AM

Q Okay. Could you read under the announcement statement that I've highlighted there for you, sir?

A Sure. There have been changes to parameter names in the National Water Information System, Parameter Code Dictionary. These changes have been 09:57AM

0042

incorporated in NWIS web. This is May 2003.

Q Okay, and when you look at, for example, on Page 2 of the exhibit, can you identify, sir, in what sense the USGS documents using the word parameter? 09:57AM

A They are using it to describe variables.

Q Using it to describe variables?

A Yes.

Q And that's exactly how Dr. Olsen used the term; correct? 09:57AM

A Well, not exactly because here the word variable isn't appearing anywhere. So apparently USGS calls them parameters, but they don't use both terms.

Q Okay. Well, Dr. Olsen used variable 09:57AM parenthetically to make sure there was an understanding that, in at least the scientific community for environmental scientists, parameters and variables mean the same thing; correct?

MS. COLLINS: Object to form. 09:58AM

A Well, I understand that that's your allegation. I don't know what was in Dr. Olsen's mind.

Q Well, isn't that also how USGS is using that term? 09:58AM

0043

A Not on this page.

Q You just testified that USGS is using the term as you would use the word variable; correct?

MR. TODD: Object to form.

A Okay. You're asking me something slightly 09:58AM different. I just indicated a minute ago that they used the word parameter to substitute for variables.

Q So do you believe that USGS is likely using the word parameter in the same way that Dr. Olsen uses the word parameter in his report? 09:58AM

A It's possible.

Q I have a question. Would you turn to Paragraph 3 of your report, sir?

A Okay. I'm sorry. Do you want me to keep this

or would you like me to give it to -- 09:59AM
Q We can just set it right here in front of you.
A Yes, sir.
Q And then sometimes we go back to previous exhibits.
A Okay, and I'm sorry, where would you like me to turn now? 09:59AM
Q Paragraph 3.
A Okay, sir.
Q Would you read Paragraph 3 for me, please?
A I'm sorry. I'm not there yet. 09:59AM

0044

Q I think it's on Page 2 of your report.
A Yes, sir. I'm currently an adjunct professor in the School of Public Health at the University of Alabama Birmingham and previously served as a professor in the business school at UAB, as a 09:59AM visiting research professor at the University of Illinois, and in other academic and professional positions.
Q So you currently hold the position as an adjunct professor at the University of Alabama; is 09:59AM that correct?
A Yes, sir.
Q You had previous positions at the University of Illinois and also another different position at University of Illinois; is that correct? 10:00AM
A Yes, sir.
Q What is the last time you taught a class at University of Alabama Birmingham?
A Well, are you talking about large classes or are you talking about dealing with graduate 10:00AM students?
Q Why don't you tell me.
A Okay. Well, three weeks ago I met with one of my doctoral students at the University of Alabama Birmingham to discuss her research in the Honduras 10:00AM

0045

regarding the transmission of tuberculosis through different sources, like textile factories versus other places.

The month before that I was working with another of my doctoral students who was finishing 10:00AM her dissertation on the impact of the caste, C-A-S-T-E, system on neonatal care in northern India, and last August I was working with Scott Keeter, who's now a professor at a university in the northeast. He was finishing his dissertation on the 10:01AM analysis of eleven different surveys for measurement of obesity.

So during the past year, I have worked with four different, four or five different doctoral students, and the last time I taught a class, a 10:01AM large class at UAB was either two or three years ago I taught the graduate level sampling theory.
Q Sampling theory?
A Uh-huh.
Q For public health? 10:02AM
A Well, it was in the department of biostatistics.
Q Okay, and so would you -- how would you characterize your current function as a professor at the University of Alabama Birmingham? 10:02AM

0046

A Well, I still have, I believe, four doctoral students working with me on different types of projects. I'm a co-PI, principal investigator, on two different research studies, one that started at UAB on the measurement and spread of obesity. The other one at Johns Hopkins University on -- dealing with issues of obesity in Hispanics, and I'm also a part-time editor for the Journal of Obesity. 10:02AM

Q Who retained you in this case?

A Actually that's a little hard to answer because I'm not sure how you would characterize them, but it was this ensemble of attorneys through the joint defense. 10:03AM

Q Who is paying your bills?

A Ozark. 10:03AM

Q Ozark?

A Ozark, Ozark Management.

Q And what is that?

A Apparently it is a company that was retained by the joint defense counsel to manage the billing process. 10:03AM

Q Have you ever worked with any of these counsel in the past?

A With Mr. Jorgensen.

Q And what cases have you worked with him in the past? 10:03AM

0047

A It was also a case involving -- excuse me -- Tyson Foods, and it was to look at a case that was filed by the U. S. Government against Tyson because of concerns about use of illegal aliens. 10:04AM

Q And what was your function in that case?

A To evaluate the work that had been done by an accounting firm for the calculation of damages.

Q Any other work with any of the lawyers in this case? 10:04AM

A No, sir.

Q Did that case for Tysons involve environmental contamination?

A No.

Q What do you know about the Illinois River watershed? 10:04AM

A What I've learned through reading the complaint and the other documents that have been in this case.

Q That have been provided to you by counsel? 10:04AM

A Yes, sir.

Q Have you ever been to the Illinois River watershed?

A I'm not sure because it's kind of a broad area, so I have to assume that I have at some point. 10:05AM

0048

Q Have you been to any location within the Illinois River watershed?

A Not that I can think of or name.

Q Have you taken any car trips, for example, that would show you the Illinois River or Lake Tenkiller? 10:05AM

A No.

Q Have you looked at any streams that might be within the Illinois River watershed?

A No. 10:05AM

Q Have you looked at any areas where there are chicken houses in the Illinois River watershed?
A No, sir.
Q Do you -- would you recognize a poultry house if you saw one? 10:05AM
A The one on my farm.
Q You have a poultry house on your farm?
A I do.
Q Do you grow poultry for commercial purposes?
A No. 10:05AM
Q So what does your poultry house on your farm look like?
A Well, it's an -- I'm amazed that the chickens don't sue me. It's a small --
Q Plenty of lawyers here in the room that could 10:06AM

0049

give them their card I guess.
A Yeah, yeah.
MR. TODD: I object.
A If there are fewer than ten chickens, is that sufficiently numerous for a class action? 10:06AM
Q So you have a poultry of maybe ten chickens?
A Yeah. It's not a big -- it's not a big combine.
Q Do you understand the size of the operations -- 10:06AM
A Oh, certainly.
Q -- of poultry growing? What size of operations do the poultry have in this particular case?
A Well, we're talking about thousands of 10:06AM chickens in a very condensed area.
Q And that's not how you grow them?
A No.
Q So when I refer to a poultry house, I'm talking about the poultry houses that are used by 10:06AM the defendants to grow their chickens.
A Specificity is important, sir.
Q Okay. So have you -- would you recognize one of those types of poultry houses if you saw it?
A I would in this because there are a couple 10:06AM

0050

down the road from my farm that are for other poultry growers.
Q Do you know what the different land use types are in the Illinois River watershed; for example, do you know the percentage of forest versus grazing 10:07AM versus urban --
A No, sir.
Q -- land use? Do you know what the potential sources of phosphorus are in the Illinois River watershed, that is, phosphorus contamination in 10:07AM ambient waters?
MR. TODD: Object to form.
A No, sir.
Q Okay. Let me make sure I restate the question because that was probably a very good objection. Do 10:07AM you know the sources of phosphorus in surface waters in the Illinois River watershed?
A Do I know all the sources or just any of the sources?
Q Have you did a study of the sources of 10:07AM phosphorus in the Illinois River watershed?

A I have not done a study.
Q What about bacterial sources in the surface waters in the Illinois River watershed; are you familiar with the bacterial sources in such waters? 10:08AM

A I have not done such a study.
MR. PAGE: Why don't we take our break here.
VIDEOGRAPHER: We are now off the Record. The time is 10:07 a.m. 10:08AM
(Following a short recess at 10:07 a.m., proceedings continued on the Record at 10:16 a.m.)
VIDEOGRAPHER: We are back on the Record. The time is 10:16 a.m. 10:16AM
Q Dr. Cowan, can we agree that when I use the term IRW, I'm referring to the Illinois River watershed?
A Yes, sir.
Q Thank you. What do you know about the hydrology of the IRW? 10:16AM
A I can't claim to have any specific knowledge of the hydrology.
Q Do you have any understanding of how contaminants move in the environment of the IRW? 10:17AM
A I haven't studied that.
Q Have you made that a study in any of your other -- in any of your other professional work?
A Well, as we discussed before, I had to study that in the toxic tort cases, at least the 10:17AM

groundwater contamination cases, because I needed to understand where the groundwater flows were occurring.
Q Okay. In this case you did not do any such evaluation? 10:17AM
A I was not asked to.
Q Okay. Was it not important to your evaluation in this case, such as similar to the toxic tort cases, for you to have an understanding of how contaminants move in the environment? 10:17AM
A No, sir. My role in this case is completely different.
Q Do you agree that the primary means of disposal of litter used in poultry production is land application? 10:17AM
A I have no opinion on that, sir.
Q So you don't have any understanding of that whatsoever?
A That's not what I said. I said I have no opinion on that. 10:18AM
Q Okay. Do you have an understanding of how poultry waste is disposed?
A No, sir, I have no opinion on that.
Q Do you have any understanding of how long poultry waste has been applied in the IRW? 10:18AM

A No, sir.
Q Do you feel like you have expertise in general concepts of fate and transport of contaminants in the environment?
A I don't know how to answer that question 10:18AM because it's so broad. Could you ask it a different

way?
Q Okay. Do you agree that rain falls within the land surface area of the IRW?
A Yes, sir, I believe rain falls. 10:18AM
Q Okay. Do you understand what the term means surface runoff?
A I have a general pedestrian understanding of the term, yes.
Q Have you ever done any professional study of 10:19AM surface runoff from lands?
A If we could combine -- confine your question to environmental studies, no, I have not.
Q Okay. Have you done any professional work on the area of infiltration of waters? 10:19AM
A I'm sorry, could you define infiltration?
Q Well, would you agree that infiltration means the downward movement of precipitation water through soil?
A Well, that's one meaning. So if that's what 10:19AM

0054

you mean, I'm willing to accept that.
Q Okay. Have you done any professional evaluation of infiltration used in that context?
A No, sir.
Q Do you -- have you done any professional 10:20AM evaluation on whether or not contaminants or constituents -- let me say this, any professional evaluation of constituents that are on or in the soils run off when rain falls on the ground in which they are located? 10:20AM
A Okay. Well, in the context that I described, my studies before, I had to look at that in the case of fertilizer plants because that's how the fertilizer got into the groundwater, but if you're asking me if I've done the environmental study of 10:20AM how that occurs, no, I have not.
Q In this case?
A Yes, sir.
Q Okay, or in that case you didn't do the environmental study of how it occurred in the other 10:20AM cases; is that true?
A I didn't do -- that is correct, I did not do the environmental study of how the runoff occurred.
Q Okay, and with regard to infiltration, would your answer be the same, that you've never done 10:21AM

0055

environmental studies of that mechanism?
A If we're confining ourselves to the definition that you gave of infiltration before, yes.
Q Okay. What would you -- would you have an alternative definition for infiltration? 10:21AM
A I've also heard the term infiltration used with regard to the merger of two flowing bodies of water.
Q Okay, and you have done evaluation of that in your professional work? 10:21AM
A Well, in the economic sense, yes.
Q Oh, in the economic sense?
A But in the environmental sense, no.
Q Okay. Some of these questions you're probably -- I just want to make sure I understand 10:21AM what you've done in this case. Did you do any independent field investigations in the IRW?

A No, sir.
Q Do you know whether any of the defendants' experts did any independent field investigations in the IRW? 10:21AM
A It's my understanding that some of them have. I know this through conference calls.
Q Did you review any for the purposes of your report? 10:22AM

0056

A No, sir.
Q Did you see any such independent investigations during the course of this case?
A I'm sorry, you mean the investigation itself or the report? 10:22AM
Q Thank you, sir. The report.
A No, sir.
Q Have you ever done any environmental investigations yourself of a large watershed?
A No, sir. 10:22AM
Q What was the purpose -- what's your understanding of the purpose for which you were retained, Dr. Cowan?
A I was asked to -- I'm going to break this into two pieces because it occurred at two different times if that's okay. 10:22AM
Q Sure.
A Okay.
Q Maybe you can give us some context of what you mean by two different times. 10:23AM
A Okay. I was initially retained to review the report submitted by Dr. Olsen, and specifically to look at the mathematics and the statistical procedures used in the analysis that he conducted.
At a later time, probably three or four months later, I was asked to look at Dr. Harwood's report. 10:23AM

0057

Q What were your objectives?
A To determine the soundness, reliability and validity of Dr. Olsen's report and then later with Dr. Harwood, the same thing. 10:24AM
Q Would you confine your objectives to the statistical analysis that was provided in Dr. Olsen's report or were you looking at the whole report?
A Well, as long as we keep in mind, so I don't get boxed in on the term statistical, the statistical encompasses everything that was done regarding the summarization of the data, the treatment of missing data, the statistical procedures used, inferences drawn, comparisons made, in other words, anything that a reasonable statistician would do to compile and then analyze the dataset. 10:24AM 10:24AM
Q And was your evaluation focused on the principal component analysis of Dr. Olsen's report? 10:24AM
A Primarily.
Q Were there other sections of his report that you also evaluated?
A Well, I looked at what Dr. Olsen said about the construction of the datasets, what he did to extract data from the larger datasets to combine data from the investigations by the plaintiffs with 10:25AM

0058

the investigations conducted by the USGS, all those other things that led into the construction of the dataset to go to the principal components analysis. 10:25AM
Q Okay. Anything else beyond the dataset and the principal component analysis in Dr. Olsen's report that you evaluated?
A And you're talking about just Dr. Olsen and not Dr. Harwood? 10:25AM
Q Yes, sir.
A Okay. Well, I just want to be sure. Not that I recall. My -- the broader scope or the scope of my analysis is contained in my report.
Q Who provided you your objectives for your work in this case? 10:25AM
A Well, I don't think anybody provided me the objectives. I was in -- in a narrow sense. In a broad sense, I was asked to read Dr. Olsen's report and critique it, critiquing it in the -- not the 10:26AM pejorative sense but in the sense of reviewing it and pointing out where it worked and where it didn't.
Q Were you told to find flaws in Dr. Olsen's report? 10:26AM

0059

A No.
Q You testified that Mr. Jorgensen retained you; is that correct?
A No, sir, I didn't.
Q Mr. Jorgensen was the first one to contact you about this case? 10:26AM
A Yes, sir.
Q Have you had most of your interface with regard to your work in this case with Mr. Jorgensen?
A I'd have to say maybe less than 50 percent. 10:27AM
Q Who was your primary contact?
A Well, Mr. Jorgensen was, but that doesn't mean that it's -- that's where most of my contact was.
Q Okay. Where was most of your contact?
A Well, it was actually through the weekly 10:27AM conference calls that we had. So I was speaking with different people at different times, and more recently I've dealt primarily with Mr. Todd, not with Mr. Jorgensen.
Q Let me hand you a document that is marked as Exhibit 3 to the deposition. Would you please identify that for the Record? 10:27AM
A Yes, sir. Okay.
Q Can you tell the court what this exhibit is?
A It is a summary of an E-mail or I'm sorry, it 10:28AM

0060

is an E-mail from Jay Jorgensen to me that follows to other E-mails back and forth, my initial E-mail from me to Jay sending my CV along and a Rule 26 document. He responded to thank me, and then he had sent me the complaint and some materials, and then 10:28AM that's followed up with another note saying here's the deposition of plaintiff's PCA expert, which I'm assuming he's referring to Dr. Olsen.
Q Could you just read the top sentence, please?
A Sure. Here's the deposition of plaintiff's 10:28AM PCA expert on his work. It's kind of long, but it shows his project and many of its flaws.
Q So you weren't directed by Mr. Jorgensen to

identify flaws in Dr. Olsen's work?
A No. This statement says that Mr. Jorgensen believes that there are flaws. It doesn't direct me to find them. 10:29AM
Q What's the date of this E-mail?
A Well, the date of this E-mail is April 27th.
Q Okay. Was that before or after Dr. Olsen submitted his expert report in this case? 10:29AM
A Well, I have to believe that it is after if there was a deposition of Dr. Olsen, but I don't actually know that for sure.
Q Did you review the deposition transcript of Dr. Olsen? 10:29AM

0061

A I did eventually.
Q How many depositions were there?
A I'm a little confused as to the number in the sense that I believe that there were two depositions and then in addition, some of the materials I was sent included testimony at the preliminary injunction. 10:30AM
Q Okay. So was this particular E-mail -- was it conveying you the deposition based on Dr. Olsen's report for the case or was it the deposition for the preliminary injunction? 10:30AM
A I don't know.
Q In the E-mail below that -- well, maybe two lines, two E-mails below the top one, it says -- I think it's from you to Mr. Jorgensen April 25th -- Jay, it was great speaking with you, and I'm really looking forward to working with you again. Is this a reference by you about the case you previously testified to? 10:30AM
A Yes, sir.
Q Have you had any previous professional experience working with agricultural issues?
A Well, yes, in several different contexts.
Q Can you identify those contexts, please? 10:31AM

0062

A Sure. I've worked quite a bit with the USDA on agricultural surveys. I was on a -- what was called a blue ribbon committee of experts to look at some of the work that was done through the USDA by one of their contractors to determine the adequacy of their quality control procedures. I worked -- while I was at the Census Bureau, some of my colleagues were responsible for agricultural surveys and censuses, so I worked with them on issues related to, for example, sample design or extrapolation from the sample to population. I've done a moderate amount of work in other countries regarding agricultural surveys, including China, Burma and a couple of Central American countries. So most of my work has been related to surveys and measurement in agricultural settings. 10:31AM 10:32AM 10:32AM
Q Were any of those surveys involving the number of animals that are grown in a particular location, like a census, census-type surveys --
A Well --
Q -- animal census surveys?
A Well, I'm hesitating because I've actually done quite a bit of work in that area but not for -- if you're talking about domestically grown -- well,

yeah, even for the domestically grown, I've done 10:33AM

0063

some of that work. I've been much more involved in determining number of animals in an area when they're not domestic but they're wild.

Q Okay. Have you done any work with the USDA concerning its animal census? 10:33AM

A Only in the sense -- well, not the US -- not the USDA, no.

Q Have you ever worked on cases involving agricultural pollutants other than what you've already testified to here today? 10:33AM

A Other than what I've testified to today, no.

Q Have you ever had an opportunity to review the census data that's produced by the USDA?

A When you use the word census, it has a very specific meaning to me. So I just want to clarify 10:34AM that if you are talking about the census data as opposed to the sample data, I've not looked at their census data.

Q Okay. What I was talking about was census data, which would, for example, tell us how many 10:34AM beef cattle are raised in Arkansas or Oklahoma or California in a particular time period.

A Okay, but it's my understanding that the Department of Agriculture doesn't always collect that information that way. They also rely on 10:34AM

0064

surveys which wouldn't include a census.

Q Okay. So I'm trying maybe -- I'm talking about the data the USDA collects concerning the number of animals produced.

A That's a better question. 10:34AM

Q That's a -- I knew we'd get to it. Have you had an opportunity to work with that type of data?

A Yes, sir.

Q Okay. Do you believe that the USDA data in that regard is reliable and valid? 10:34AM

A Well, those are two different mathematical concepts.

Q Okay. Let me separate them out then.

A Okay.

Q Do you believe they are reliable? 10:35AM

A In most cases, within the sort of restrictions that you would ordinarily have with a federal survey with a limited budget, yes.

Q Okay, and what about validity; do you believe they are valid? 10:35AM

A Yes.

Q Do you believe they're -- let me restate that. Strike that. Do you believe that the data that USDA provides concerning number of animals can be reasonably relied upon in order to understand the 10:35AM

0065

production of a type of animal from a certain area during a particular time period?

A I don't know how to answer your question because it would depend on the area and depend on the time period. 10:35AM

Q What about for Arkansas and Oklahoma?

A Probably at a state level, the numbers would be pretty good on a five-year basis.

Q What about a county level?

A Probably less reliable. 10:36AM
Q Would you use them to determine the amount of animals that are -- that exist in a particular county; would you rely upon that data for your analysis?
A I don't know because I don't know what has been done with the data. I mean, you're describing -- you're asking such a broad question that I can't answer it because there are lots of different ways to come up with county estimates from a state level survey. 10:36AM
Q Well, if I'm looking at the amount of poultry that was produced in Washington and Benton Counties of Arkansas over the last ten years, would you rely on the USDA data for those numbers?
A Well, I'm going to define rely for you. I 10:37AM

0066

would start with the USDA data and then determine what the estimation procedures were that were used applied to that data to estimate for the two counties you said for a ten-year period, and let me also say that it's entirely possible, given the way 10:37AM the USDA does sampling, that they have perfectly reasonable samples within those two counties, but I wouldn't know without going to the USDA and reading through everything that the USDA publishes to figure out whether they're sampling in those two specific 10:37AM counties or if you're talking about census data or if you're talking about a procedure called borrowed strength or James Stein estimation, James Stein, S-T-E-I-N.
Q If -- 10:37AM
A Excuse me, I'm not done. That would allow me to then make the estimate. So I'm having trouble with your questions because they're so broad and there's so many different possibilities.
Q If you let me interrupt, maybe I can narrow 10:38AM it.
A Thank you. I was done.
Q Have you done any evaluation of the USDA animal production data in Arkansas or Oklahoma with regard to poultry, cattle, swine? 10:38AM

0067

A No, sir.
Q Have you ever used principal component analysis in your professional work?
A Yes, sir.
Q Could you explain to me in general terms the 10:38AM applications in which you've used principal component analysis?
A Sure. Do you want a short list or the full list?
Q Could you kind of categorize how you used it? 10:38AM
A Sure. Remember earlier we were talking about my graduate students?
Q Yes.
A My most recent graduate student is using principal components analysis on a survey conducted 10:38AM in Honduras to look at -- she's conducting a behavioral analysis to determine whether she can find ways to help workers stem the flow of multidrug resistant Tuberculosis in the Honduras.
Q I'm going to apologize for interrupting. 10:39AM

A You bet.
Q But can you tell me applications where you used PCA in your own work, not maybe working with someone else? For example, have you done any studies yourself where you've used principal 10:39AM

0068

component analysis?
A Okay. Just to conclude what I was saying, however, I will say that I consider that my own work. I'm advising a doctoral student, but if you're asking me if I've done the work as opposed to 10:39AM working with somebody else, actually the very first work that I did was for the National Science Foundation doing an analysis of economic data for a country to determine sort of sources and flows of income and how the economy within that country 10:40AM operated, somewhat like the structure of our own national income accounts.
Since then I've used principal components, for example, in a -- in studies of samples of people to determine whether or not you could use principal 10:40AM components and its adverse Mahalanobis distances for sampling purposes for construction of samples using controlled selection. I've used it in a financial context where we've looked at, for example, stock data. You've got lots of different types of stocks, 10:40AM and the question is if you are trying to invest in stocks, how do you classify them together or apart and is there a more efficient way to classify stocks relative to other methods of creating equity within a firm? Those types of analyses are to determine 10:41AM

0069

the structure of financial markets. So a lot of different applications.
Q So has your work in the -- with PCA been primarily involving studies within the social sciences? 10:41AM
A Yes.
Q Okay. Have you ever done any work with PCA in the non-social sciences?
A That seems so harsh. We could call them less social. 10:41AM
Q How would you call it?
A I understand what you meant. What are commonly referred to as the hard sciences.
Q Yes, sir.
A Well, only in the sense of deal with it from, 10:42AM you know, pure mathematical, which really isn't the social sciences, but if you're talking about like physics, chemistry and so on, no.
Q Or geochemistry?
A No. 10:42AM
Q What about samples involving environmental contaminants?
A Could you be a little bit more explicit?
Q Well, like in this case where Dr. Olsen was reviewing samples of -- environmental samples and 10:42AM

0070

testing it for different parameters, geochemical parameters; correct?
A Uh-huh.
Q Have you done any kind of PC analysis with a dataset similar to Dr. Olsen's? 10:42AM

A No.
Q Have you published any peer-reviewed articles concerning principal component analysis, whether it's the social or hard sciences?
A Well, there was a report to the National Science Foundation. So they published it, I didn't publish it, although that was a really long time ago, and then there are two papers in my resumT that are -- describe the use of Mahalanobis distances, which is the adverse of principal components, for essentially attempting to do controlled selection -- use of controlled selection methods in sample surveys. 10:42AM 10:43AM
Q And what kind of survey was involved; was it a social sciences survey?
A No. This was for the Bureau of the Census. So it would be in general any of the surveys that they do. 10:43AM
Q People population surveys?
A No, sir. At least half or more of the work that's done by the Census Bureau is business surveys, surveys of governments, surveys of farms. So surveys on almost anything, but not necessarily people surveys. 10:43AM
Q Okay. I'm sorry. My poor choice of words, but those -- none of those studies involved the data -- hard science data; correct? 10:44AM
A Not the way we were discussing hard science before, no.
Q Did you have to study PCA applications for -- when I say -- maybe it would make it easier if I kind of define environmental sciences. I'm talking about an environment case like we have here. 10:44AM
A Okay.
Q A contamination case. So when I say that, I'm not talking about maybe my sociological environmental, the way I grew up or something like that. I'm talking about contamination-type cases; okay? 10:44AM
A Uh-huh.
Q Did you have to study the use of PCA in environmental analysis before you did the work in this case? 10:44AM
A No.
Q And why not?

A Well, you asked me if I had to study the use of PCA in environmental cases, and I took your have to meaning it was an absolute must to be able to understand PCA. PCA is a common technique that's been used for a very long time, and I've used it throughout my career. So if you're asking me if I had to study PCA, no. 10:45AM
Q Okay. Let me ask you this then: Would you agree that the application of PCA to environmental sciences is somewhat different than when you apply it to the work you've done in the social sciences? 10:45AM
A No.
Q You say it's the same methodology?
A Well, mathematically, the mathematics aren't going to change. 10:45AM
Q You don't think there's any unique attributes

of doing work in environmental science data that would be important for you to appreciate prior to evaluating Dr. Olsen's work in this case?
A Well, let me put it in perspective. What Dr. Olsen did was he did his analysis using a program called SysStat, which is one of the programs we use, and SysStat doesn't ask if it's environmental. It just runs the program. 10:46AM
Q Okay, and you're --

A So the mathematic -- I apologize because I interrupted you, but just I wanted to conclude by saying the mathematics are exactly the same.
Q Okay, but in your use of PCA, isn't it important to have an understanding of the types of data that are involved in the PCA analysis in order to interpret that data? 10:46AM
A Well, that's why we reconstructed all of Dr. Olsen's datasets.
Q But did you come to an evaluation and understanding of the type of data that was involved? 10:46AM
A Well, I came to some understanding of the type of data. I'm not putting forth -- myself forth as a chemist, a biologist or anything else, but, you know, when I work with doctors and I design research for them, I'm not putting myself forth as a physician, but that doesn't mean that my work isn't, you know, valuable in terms of understanding the transmission of diseases. 10:46AM
Q Did you do any additional study of PCA applications in environmental forensics prior to doing your work in this case? 10:47AM
A I did.
Q And what did you do?
A Well, I'm sorry. I'd like to amend just the 10:47AM

word prior. I did it concurrently.
Q Okay, and what did you do?
A Well, I -- first, I read Dr. Johnson's chapter in the book that he published. I also read a text book by a Professor Jeliffe, J-E-L-I-F-F-E I believe, that has a couple of chapters on use of PCA in environmental work. I looked at other articles that had been referenced that use PCA in environmental work, and I believe that there's an example also given in geology, another of the hard sciences, not in Harmon's textbook but in a third textbook I have and, I'm sorry, I can't remember the name of that one. 10:47AM 10:48AM
Q And why did you do that review and evaluation?
A Just to understand what other -- to put the analysis in context and understand what was commonly done in that field as opposed to my field. 10:48AM
Q Did you find that to be important in review of PCA analysis?
A No, sir. 10:48AM
Q So you don't think it was important to know what the common practices are, for example, in the environmental science field as opposed to your field in order to understand whether the environmental scientists properly employed PCA? 10:48AM

A From a mathematical perspective, what I found

in reviewing -- you asked me if it was important. The reason it wasn't important was because I didn't learn anything new in reading those articles or the journals or the books that I didn't already know in terms of the mathematical application. So it couldn't be important if I wasn't learning something new or different. There wasn't anything different.

Q Do you have any experience prior to this case in transforming environmental sampling data? 10:49AM

A Remember the barge case we were discussing before?

Q Okay.

A I had to do transformations on that data and deal with some of -- well, I had to do 10:49AM transformations on that data.

Q What kind of transformations did you use?

A Some cases logarithmic and other cases calculation of logistic values, which is -- uses a log but there's a further set of transformations 10:50AM involved.

Q Was it a Log10 transformation?

A I believe it was, yes.

Q Okay, and why did you do the transformation in that particular case? 10:50AM

0076

A Excuse me. Well, for a couple of reasons. One was that I was attempting to smooth out the data and determine whether I had outliers. Smoothing the data has to do with essentially putting it into a linear form as opposed to a curval linear form. 10:50AM That was the primary reason.

Q Was it -- is it fair to say you did that transformation to avoid having to use a skewed dataset?

A No. 10:51AM

Q No?

A Well, I'm sorry. I think of the word skewed differently than you do. So could you define what you mean by skewed?

Q Well, that's to -- like you mentioned those 10:51AM outliers -- to normalize the dataset so you could have -- so you wouldn't have an outlier unfairly influence your evaluation of the dataset.

A There -- well, first of all, thank you for the clarification. In my earlier response, I was 10:51AM thinking of skewed in a different context that's probabilistic that has to do with dis -- probability distributions have characteristics of central tendency skewness and kurtosis. So my answer before wasn't relevant to what you were asking me. 10:51AM

0077

To answer the question that you just asked me, the -- I would use logarithmic transformations to put -- if you are trying to run a regression, the whole purpose of the regression or the principal components, which is another type of linear 10:52AM transform, is to have things actually in a straight line, but to deal with outliers, there are other procedures. First of all, you need to detect if there are outliers and then determine whether or not the outliers are true outliers in the sense that 10:52AM they're mistakes as opposed to being unusual values but correct, and then, finally, you also want to

know from the regression or principal component perspective whether or not the outliers have significant leverage. 10:52AM

Q Was all the work that you did in this case reflected in your report, Exhibit 1?

A Well, I don't know how to answer that question either because there were five of us working, and so there is other work that we did that helped build up the data or to make comparisons but isn't necessarily reflected in the report. There's a limit as to how much you can put into the report. 10:53AM

Q Okay. Well, first of all, tell me who the five you are referring to are. 10:53AM

0078

A Okay. Within my office there is Dr. Ed Reeves. There's a gentleman Mauricio Vidaurre, V-I-D-A-U-R-R-E, and you're own your own for Mauricio, and within my Birmingham office there is Danny Heisner, H-E-I-S-E -- H-E-I-S-N-E-R, and Marcellus Smith, S-M-I-T-H. 10:53AM

Q And what -- did any of these people perform the underlying analysis that supports your work reflected in your report?

A Well, we didn't actually work that way. Dr. Reeves and Mr. Vidaurre both performed various analyses, but then I redid the analyses to be sure that they were correct or if they were not -- if we didn't match, then I found out why. So I did the analyses that underlie what's here, but they also did analyses at my direction, and then if they -- after I saw their outputs, I had them give me back the data, and then I ran the analysis independently to see if I got the same result. 10:54AM

Q And the other two individuals, what function did they perform, that is, Smith and Heisner? 10:54AM

A They are data processing people, so they took data. Well, the first thing they did was they downloaded everything from the Exponent database that was maintained by the defendants. They 10:55AM

0079

downloaded anything -- well, they downloaded everything initially because it was easier to just grab everything, and then they separated out the data that was in there from different datasets and provided -- well, they -- and then they provided the data directly to Dr. Reeves. 10:55AM

Q So the work that -- the materials you got in this case came from Exponent?

A As -- I believe that's the correct name. There's a large FTP site where you could go and look at all the data that had been received, and it was uploaded into a common dataset. 10:55AM

Q Did any of these people you just mentioned write any of the portions of your report?

A No. 10:55AM

Q Now I want to go back to my original question. Was there any analysis that you did when you were evaluating Dr. Olsen's work that was not put into your report?

A I think my answer still stands. There are other things that we did analytically that were supportive but aren't necessarily in the report. 10:56AM

Q Are there any categories of analysis that are

not reflected in your report?
A I'm not sure what you mean by -- 10:56AM

Q You have certain general conclusions and opinions in your report, do you not?
A Yes, sir.
Q Okay. Did you come to any conclusions or opinions that you did in your work in this case that 10:56AM are not reflected in your report?
A Not that I can recall at this time.
Q Were you informed that you were to retain all materials received or reviewed for the work you did in this case? 10:57AM
A I don't remember if I was informed of that, but that's pretty standard practice for us anyway.
Q So you did so --
A Yes, sir.
Q -- in this case? And did you produce all 10:57AM those materials to your -- the lawyers in this case?
A Yes, I did.
Q Who did you produce them to?
A Well, I produced them at two different times. So I produced some materials to Mr. Jorgensen, and 10:57AM then later I was asked to provide everything that we had downloaded from Exponent that had anything to do with Dr. Olsen, just the rough-outs, not what we did with them, and since we had created a separate partition in a server, it was easy for us to do that 10:57AM

and put those on to DVDs, and we provided those to Miss Southerland.
Q When was that?
A The latter or the former?
Q The latter. 10:58AM
A I believe it was about three weeks ago, four weeks ago. I'm sorry, I don't remember exactly.
Q And the materials you provided to Miss Southerland three weeks ago were the materials you reviewed from Dr. Olsen's files; is that a correct 10:58AM characterization of what you just --
A Well, not exactly, but I'll explain why. It was everything we downloaded from Dr. Olsen's file, but since there was a lot of materials in Dr. Olsen's files or at least the files that were on 10:58AM Exponent that had nothing to do with the case, I didn't review it.
Q Okay. So what did you -- so are you saying that the materials you provided three weeks ago were materials that you did not review in this case? 10:58AM
A That's exactly what I didn't say. Let's try again.
Q Yeah. I'm trying to understand what you produced three weeks ago and whether or not it was part of materials you considered in this case. 10:59AM

A Okay. I understand that's what you're asking but you mischaracterized my testimony. What I said was, we did a download of everything off of the Exponent database that was in the Olsen files, okay, put it into a separate partition on one of our hard 10:59AM drives, okay, and then we extracted out data and we extracted out reports, but there's a considerable amount of material that Dr. Olsen provided that has

nothing to do with the case. I didn't review
anything that wasn't relevant to the case. For 10:59AM
example, he provided copies of his auto exec bat
files, which are necessary for running his computer,
but I don't think have anything to do with the case.
Q May I interrupt you? I'm just trying to find
out what you produced three weeks ago. 11:00AM
A And that's what I'm telling you.
Q It's stuff you did not --
A I produced --
Q It's stuff you did review in the case for your
opinion or not? 11:00AM
A Both.
Q It's both, and was that material produced in
the early production or was this a group of new
material that wasn't previously produced?
A Well, what we provided before was anything 11:00AM

0083

that was relevant to the case, in other words, all
the datasets, all the reports, but not Dr. Olsen's
auto exec bat files, and we produced that in our
first production.
Q So was there -- was there a duplicate of 11:00AM
materials in the first and second production?
A I believe so.
Q Okay. In your first production did you
produce all the materials that you reviewed and
considered for your opinions in this case? 11:00AM
A I did.
Q Okay, and what was the purpose of the second
production then?
A I was asked, and I don't know why, but I was
asked to provide everything that we downloaded off 11:00AM
of the Exponent files for --
Q Whether you reviewed it or not?
A Whether we reviewed it or not.
Q Did you produce all your E-mails --
A Yes, sir. 11:01AM
Q -- with counsel?
A Well, I produced all my E-mails, so that would
be with anybody.
Q With anybody in this case, okay. What about
the materials that you reviewed -- you had regular 11:01AM

0084

conference calls; is that correct?
A Yes, sir.
Q And you had an FTP site that you reviewed
materials on an FTP site; is that correct?
A There was an FTP site. On occasion I would 11:01AM
review materials on the FTP site, but that was sort
of inconvenient.
Q Okay. When -- did you produce all the
materials that you reviewed on these FTP site calls,
that is, if there was something that was provided 11:01AM
over the FTP site, did you review and during the
call did you produce the image that was provided or
the information that was provided?
A Well, the nature of that was that I didn't
have access to it. That's the whole idea behind 11:02AM
having an FTP site. It was on my screen, but at
that point WebEx or whatever program it was we were
using had taken over my computer, so I didn't have
access to those materials. I could see them but

they weren't retained. 11:02AM
Q Okay. So you did not produce the materials you were shown over the FTP site?
A I was not capable of producing those materials.
Q The answer is yes or no. Did you produce them 11:02AM

0085

or not?
A No, I did not.
Q Did you ever receive, after you saw the materials on the FTP site, those materials in another format? 11:03AM
A Once.
Q Did you produce those materials?
A I did.
Q And what was that?
A Dr. Johnson's report. 11:03AM
Q Do you recall what other materials were shown you on the FTP site?
A Mostly maps.
Q Maps?
A Maps. 11:03AM
Q Were they involved with analysis performed by experts in this case?
A Yes.
Q Were they maps of locations of information within the IRW? 11:03AM
A For the most part, yes.
Q Do you recall what the maps were when they weren't referencing the IRW?
A Well, there's some population data for cities that I guess technically would be in the IRW, but 11:04AM

0086

they are not part of -- they're cities; they're not the IRW.
Q How often did you have these web meetings?
A Typically every Thursday.
Q From when to when? 11:04AM
A Probably -- I don't remember exactly when they started. I believe it was around June or July, and they continue to this day, although sporadically.
Q Did you have any face-to-face meetings?
A With -- 11:04AM
Q With the experts in this case?
A No.
Q Did you have any face-to-face meetings with counsel in this case?
A Just with Mr. Jorgensen. 11:04AM
Q And when was that?
A I was in Washington, D.C. on another matter, so I called Jay and asked if he would like to have lunch.
Q Any other meetings with counsel? 11:05AM
A Well, there was the initial meeting before I was retained.
Q Any other meetings with counsel?
A Not that I recall.
Q So the meetings with experts in this case were 11:05AM

0087

primarily these web meetings that were hosted on an FTP site?
A They were only those meetings.
Q Only those meetings. Which of the other

experts for the defendants in this case, their reports, have you reviewed in this case? 11:05AM

A The experts for the defendants?

Q Yes, sir.

A Okay. Dr. Johnson -- Dr. Johnson's report I reviewed, and then I had discussions with but didn't review their reports with -- I believe it was Mr. Clay. 11:05AM

Q Any other reports or opinions you discussed with other experts for the defendants in this case?

A I had an exchange with -- was it Mr. Nobles? I -- at this point it's a while ago so I'm not going to remember the names exactly -- just to discuss population estimates. 11:06AM

Q Did you have any discussions with a Dr. Murphy about his opinion in this case? 11:06AM

A I wasn't aware that there was a Dr. Murphy in this case until yesterday.

Q I guess the answer is no; the answer is probably no?

A No. 11:06AM

0088

Q Okay. Tell me about did you review a previous version of Dr. Johnson's report or did you review the final version that was submitted to the plaintiffs in this case?

A I submitted -- what I reviewed was the version that was on the WebEx. 11:07AM

Q So you actually got an actual copy of that --

A No, sir.

Q -- by another form?

A Well, it was given page by page on the WebEx and you could read it on your screen. 11:07AM

Q Oh, I see. You never received Dr. Johnson's report other than on the WebEx?

A Well, after it was submitted --

Q Right. 11:07AM

A -- then I did.

Q This was a draft -- you understand that what you saw was a draft of Dr. Johnson's report before it was finalized?

A Yes, sir. 11:07AM

Q And when was that?

A Probably -- I'm sorry, I'm not remembering dates exactly, but it would have been in late October or early November approximately.

Q Were you the only person reviewing the WebEx 11:07AM

0089

viewing of Dr. Johnson's report?

A I believe there were one or two other experts on the phone call.

Q Okay, and what was the purpose of the call?

A As I understood it, to review and provide commentary and ask for clarification in his report. 11:08AM

Q Okay. Did you provide any comments at that time?

A On the phone, yes.

Q Okay. What were your comments? 11:08AM

A Well, most of the discussion I had with Dr. Johnson had to do with two things. One was asking for clarifications of certain things that he presented in his report because I felt that the average reader would not be able to understand the 11:08AM

point he was trying to make, and so it had to do with presentation and how understandable it was to a non-technical reader.
Q Anything else, any other types of comments?
A The other types of comments had to do with the differences or more the overlap between his findings and my findings, since he also looked at the use of PCA. 11:09AM
Q Did you find any inconsistencies between Dr. Johnson's findings and your findings? 11:09AM

A Not that I recall.
Q Okay. So what were the -- could you give me example of a comment you made on the overlap portions of Dr. Johnson's report and your report?
A Sure. Well, there was some discussion about the treatment of missing data and the impact of missing data on the overall analysis and in particular, differences between analyses between SW3 and SW15, with one dataset having no missing data and the other having some missing data. He would -- was interested in that, too, and the question was the impact on some of the analysis. We also -- 11:09AM 11:09AM
Q Did you come to any conclusions on that matter?
A Yes. They're presented in my report. 11:10AM
Q Okay, and what is your conclusion?
A Well, that there's a very large volume of missing data, and the way that the missing data was treated for principal components analysis introduced biases into the analysis. 11:10AM
Q So you felt the missing data had a substantial impact on the results of the PCA; is that correct?
A Yes, sir.
Q Okay. Did Dr. Olsen agree with that -- excuse me, did Dr. Johnson agree with that? 11:10AM

A Yes, he did.
Q Okay. Anything else you discussed about the overlaps?
A Dr. -- actually Dr. Johnson also discussed something that's related to the missing data which has to do with the non-detects, and we talked about types of tests, what non-detect levels would be appropriate or recorded, and that was informative for me because of the impact in the analysis I was already doing on logarithms. 11:10AM 11:11AM
Q Well, with regard to non-detects, did you and Dr. Johnson have similar critiques of Dr. Olsen's report on how he treated non-detects?
A We did after we had a conversation.
Q So it's your view that Dr. Johnson has critiqued Dr. Olsen on his treatment of use of non-detects? 11:11AM
MR. TODD: Object to form.
A I don't recall seeing that in Dr. Johnson's report. What I said was we discussed it. 11:11AM
Q And he agreed on the phone but didn't put it in his report?
A I just don't remember.
Q Anything else?
A We both talked about the retention of the 11:11AM

number of principal components because both of us had views as to the number of principal components that were coming out of the analysis being conducted.

Q Okay, and did you guys have an agreement on your concerns in that regard? 11:12AM

A Yes. Both of us felt that the number of principal components retained was too few, and that both of us pretty much used the same testing methods to determine the appropriate number of principal components. 11:12AM

Q How many principal components did Dr. Olsen review?

A Well, what he presented in his report and talked about were two. 11:12AM

Q How many did you review --

A Well --

Q -- and do an analysis on?

A Well, he must have in his initial runs out of SysStat, if he had taken all the defaults, he would have gotten five. 11:12AM

Q Only if he had taken all the defaults?

A Well, not only if he had taken all the defaults. There are other ways to get five, but that would be the common number. 11:13AM

0093

Q Is it your understanding that Dr. Olsen did in fact review five different principal components in his analysis of PCA?

A I don't recall that he reviewed five. I know he had the materials available to review five, but I don't know that he actually reviewed five. 11:13AM

Q Nothing in his report would indicate to you that he considered five principal components?

A Well, in the report, as I recalled, there are ancillary statistics that would tell you the number of principal components to review. Whether or not he actually looked at those or just discarded them immediately, I don't know, but, for example, scree plots would tell you that you would take five. The listing the Eigenvalues which Dr. Olsen provided would tell you to take five. 11:13AM

Q Okay.

A So there are other -- other measures that would tell you to take that number. I just -- but I can't speak to what Dr. Olsen did or didn't do. 11:14AM

Q I digressed. Back to the FTP discussion with Dr. Olsen and his reports on the screen --

A You mean Dr. Johnson?

MR. TODD: Dr. Johnson?

Q Excuse me. Dr. Johnson? 11:14AM

0094

A Yes, sir.

Q Could you tell me anything else you discussed concerning his report on these areas of overlap?

A I don't recall, but since we're about to take a break, I would be happy to ponder that and answer your question after the break. 11:14AM

Q Let's do that then.

A Okay. Because I don't -- nothing jumps out at me.

Q Thank you. We'll take our break. 11:15AM

VIDEOGRAPHER: We're now off the Record.

The time is 11:14 a.m.
(Following a short recess at 11:14 a.m., proceedings continued on the Record at 11:31 a.m.) 11:31AM
VIDEOGRAPHER: We are back on the Record.
The time is 11:31 a.m.
Q Dr. Cowan, your counsel just told me that before we go back to the question that's pending before our break, you had something you wanted to 11:32AM clarify from your previous testimony.
A Yes. When you asked me if I met with counsel before, I took before to mean before this week, and I didn't include yesterday's meeting with Gordon and Melissa. So I did meet with them yesterday, too. 11:32AM

That's the only clarification.
Q Thank you.
A Okay.
Q Now, did you have a chance to think of any other discussions you might have had when you were 11:32AM reviewing Dr. Johnson's draft report?
A The only other real discussion that we had with respect to his report was on the use of rotations to interpret the results.
Q And what was that discussion? 11:32AM
A Well, pretty much that in the work that we were doing to review Dr. Olsen's work, we were using rotations to get an interpretation of what some of the factors were, and I asked Dr. Johnson if that was routinely his practice, and he said yes. So it 11:33AM was a pretty brief discussion; that was it.
Q Do you know whether or not Dr. Olsen did review various rotations when he was interpreting his work on the PCA?
A Yes. He says so in his report. 11:33AM
Q Did Dr. Johnson make any changes in his report based on discussions with you?
A Well, I know he did respond to the clarifications that I suggested so that things were written in plainer English, maybe a little bit more 11:33AM

expansive to cover whatever it was.
Q What about the second type of discussions where you talked about overlaps; did he make any changes in his reports based on these overlap discussions you just testified to? 11:34AM
A No, sir, because when we discussed it, we discussed that although we were overlapping, that it was probably just fine, that redundancy in some cases is a good thing.
Q Did you make any alterations to your report 11:34AM based on any of your discussions with Dr. Johnson?
A The only -- I didn't really make changes. I acquired a better understanding from Dr. Johnson about the multiplicity of tests that could be performed and one -- what non-detect levels -- or 11:34AM what non-detect levels I would likely see. So I gained a better understanding of non-detect levels from speaking with Dr. Johnson.
Q Did you not understand what a non-detect meant in -- 11:34AM
A Oh, no.
Q -- environmental data before your discussions

with Dr. Johnson?
A I apologize for interrupting you. No. I understood perfectly what a non-detect was because we see that frequently in biostatistics, too. What I wanted to know about was the specific tests and what a non-detect level would be for different tests of the same analyte. 11:35AM
Q Okay, and did you -- what did you learn new from Dr. Johnson in your discussion on non-detects? 11:35AM
A That there were different levels of precision for different types of tests for the same analyte, and that sometimes there was a preferred test and sometimes there wasn't. 11:35AM
Q I think we're going to come back to that subject in a minute, but before we go there, could you summarize for me today what your opinions are that are contained in your report?
A Certainly. May I refer to my report? 11:35AM
Q Yes, sir.
A Thank you.
Q What I'm trying to do is understand what your key opinions are in the case.
A Okay. I realize there was one other change that I made to the report after speaking with Dr. Johnson, and that was simply that I expanded the section that I had on strength of relationship. 11:36AM
Q Can you be a little more specific what you mean by strength of relationship? 11:37AM
A If you think about two variables being correlated, that's a concept that I understand because I work with it every day, but I find that I have -- when I speak to somebody who isn't trained in standards or who doesn't work with this every day, that they have trouble developing a gut-level reaction or understanding of what it means for two variables to be correlated, and so I expanded my section to give examples of strength of relationships so that the reader could understand that a perfect correlation would essentially just be a straight line. 11:37AM 11:38AM
Q What section are you referring to now?
A It is --
Q Can you give me your paragraphs, please? 11:38AM
A Sure. Okay. Starting with an introductory Paragraph 21, there was an entire section labeled Strength of Relationship that continues on Pages 10 and 11, Paragraphs 22 through 26, and includes Page 12 also. 11:38AM
Q Okay, and so you modified this to make it clearer after you talked to Dr. Johnson; is that your testimony?
A Yes.
Q Okay. Now, I think the question before you is, would you give us an overview of the opinions you plan to provide in this case? 11:38AM
A Yes, sir, and I apologize for the interruption. Okay. My first observation about the work done by Dr. Olsen is that you have data from a number of different sources and you also have different numbers of observations on different 11:39AM

variables. So I get back to the sources later, but to expand on the problem with having different numbers of samples, for most of the chemicals, like copper or iron, you only have one measurement per sample, but for bacteria, you can have a variable number, like four or three, and so because of that, what Dr. Olsen did in his analysis was he took the average of those bacterial values, and in doing so, you essentially create a record that has only one observation but the original data had multiple observations on bacteria, which means that bacteria is going to be more variable in some sense. Any of the four bacteria measures that Dr. Olsen used are going to be more variable than the other measurements, but that's not reflected in the principal components, and since principal components is a tool used to summarize or to explain variance, if you're not reflecting the variability in the full sense, sampling and modeling variability, then the principal components may be -- the outcomes might have been different.

Q So your complaint -- one of your first complaints is that Dr. Olsen averaged multiple observations?

MR. TODD: Object to form.

A No, I'm not objecting to the averaging. The averaging is fine. What I'm objecting to is that he didn't reflect the variability with those observations because the averaging is just a central point, but, in fact, the bacteria had a great deal of variability in the individual observations, and so I'm not objecting to the averaging. I'm objecting to the use of the averages in the principal components when the principal components are designed to study variability.

Q Can you show me in your report where you introduce this criticism?

A It is all of Page 16 and 17, Paragraphs 36 through 41.

Q Okay. What's next?

A We went back to the original data that had been provided by Dr. Olsen and we obtained through the Exponent database, and we then tried to reconstruct the Excel databases that Dr. Olsen used in his analyses. The Excel databases were the input to the SysStat programs, and the problem was if you started with the original data, you couldn't get to Dr. Olsen's Excel databases.

Q Was the original data Access database?

A Yes.

Q Were you the one that tried to access the Access database to get the data for the Excel spreadsheet to go into the PCA analysis?

A No. My staff did.

Q Okay. So you don't know whether or not they were able to find the query file in the Access database, do you?

A Well, I worked with them, and there are multiple query files. So are you referring to a specific query file?

Q Well, if I look at the query file in the

Access database that would allow you to query the information to get the SW3 data. 11:43AM
A As near as I can tell, from looking at their work, no, they didn't find one.
Q What's the next item that you critiqued Dr. Olsen, big picture?
A Okay. There's a lot of missing data, and the 11:43AM

0102

problem with -- there are several problems with the missing data, and it's not that the datasets don't have missing data. It's the use of the or the methods used to account for the missing data. So one problem is the volume of missing data because it 11:44AM is a very large volume of missing data, especially for bacteria, and I note in my report only 47 percent of the observations that were analyzed are complete in the sense that they have observations on every single variable. That means that 53 percent 11:44AM had one or more values that were substituted.
Q Okay, and where in your report is this criticism --
A Page --
Q -- introduced? 11:44AM
A I'm sorry. Page 19.
Q Is that paragraph --
A 44, 45 and 46.
Q Okay. What else? Were you finished discussing that, at least in a summary form, sir? 11:44AM
A Thank you for asking. I just want to take a quick look here. No. I'm going to come back to these later.
Q All right. What's the next item? I guess we're up to Item No. 4? 11:45AM

0103

A Item No. 4 is the fact that data came from two large sources. One was the sampling done by the plaintiffs and a second source is the U. S. Geological Survey, and they were combined, but there's no testing or no discussion about 11:45AM differences between the two datasets.
Q Okay. So you're critical of combining the CDM testing with the USGS testing that was done?
A I'm not actually critical of combining them. I'm critical of combining them and then not 11:45AM determining what the impact was and whether that added to the variability in the data -- the variability that wasn't accounted for.
Q Do you know whether or not Dr. Olsen did such an evaluation? 11:46AM
A It's not reported in his report or at least I don't recall it being there.
Q Okay, and where in the report is this issue introduced, sir?
A It is on Page 20, paragraph -- I'm sorry, 11:46AM Pages 19 and 20, Paragraphs 46 and 47.
Q Okay. What's the fifth issue, if there is a fifth?
A Well, as I described it before, there are two -- you can divide the data into two groups. One 11:46AM

0104

set of observations is the set of observations that has no missing data at all. So it's a complete dataset, and then the second larger set of

observations would be the observations that have some missing data. It could be missing one variable or could be missing multiple variables. In looking at the observations that you have with and without the missing data, so now I'm talking about the real values as opposed to the values substituted because they were missing, and -- 11:47AM

Q Is it your criticism that -- sorry to interrupt but I want to make sure I understand that. Is it your belief that Dr. Olsen substituted for missing values in his PCA analysis?

A He says he does, so, yes. 11:47AM

Q How does he substitute for them?

A He uses the mean value for each variable and substitutes that, and I have a discussion of that later.

Q Okay. So go ahead now with the data with missing -- with and without missing data. 11:48AM

A Okay. Ignoring the missing data but dividing the dataset into two groups, one that has complete data and a second one that has incomplete data, if you compare the values on the variables for those 11:48AM

0105

two groups, there are -- on many of the 26 variables that Dr. Olsen analyzed, there are huge differences that you wouldn't expect if the data was just missing at random, which is one of the assumptions. The missing at random is one of the assumptions 11:48AM commonly made in conducting any sort of statistical analysis.

Q And where is this criticism first introduced and discussed in your report?

A It starts with Paragraph 48 on Page 20, 11:48AM continues with Paragraph 49 and Paragraph 50 and 51.

Q Does it include also Chart 6 in your report?

A Yes.

Q Okay. What else; what other key criticisms do you identify? 11:49AM

A Dr. Olsen, as we were just discussing, substitutes means any time that there's a missing value. This changes the distributional properties of the data and in particular, by substituting the means when you've got differential -- differential 11:49AM amounts of missing data, it biases the results.

Q And is that introduced, sir, in Page 22, Paragraph 52?

A Yes, and then continuing on, including the charts through Paragraph 56 on Page 25. 11:50AM

0106

Q I notice you don't state in the report where Dr. Olsen states that he actually substituted the means for missing data.

A Well, I don't recall where that is, but if you go back through the calculations from the original 11:50AM dataset through complete replication of Dr. Olsen's results, that's the only way to get there. Plus, if you look at the data from the beginning to the end, you can look and see that those are the means that are substituted. 11:50AM

Q So you assumed he substituted missing data; Dr. Olsen never stated that in his report; correct?

A I just don't remember if Dr. Olsen stated it in his report, but I didn't have to assume it. I

could look it up for myself. 11:51AM
Q It's your opinion that if there's missing data in a particular sample, you can't run SysStat on it, can you?
A No, I didn't say that.
Q Okay. Well, then how would you get to the conclusion that Dr. Olsen had to substitute mean values for missing parameters in his samples? 11:51AM
A It's in his dataset. So in looking at his datasets, the values are present there.
Q The mean values are present there? 11:51AM

0107

A That's my recollection.
Q Okay. What else?
A I wanted to make one more observation. To be able to do the final calculations that Dr. Olsen did to generate the PC scores, he couldn't have 11:52AM generated all of the PC scores without plugging in the means. So it just wouldn't be mathematically possible.
Q So that's also your testimony, that Dr. Olsen substituted the means of the -- of the -- of that 11:52AM particular analyte when he calculated the PC scores?
A Yes.
Q Okay. If you are mistaken and in fact Dr. Olsen did not substitute the means for any missing value when he did his PCA, would that affect your 11:52AM criticism of this aspect of his report?
A Well, are you asking me if I'm mistaken and it wasn't Dr. Olsen, it was SysStat that did it, or are you asking me if I'm wrong and it wasn't the means?
Q I'm just -- I'm asking you -- are you 11:53AM suggesting that if there's a missing value, SysStat will substitute the mean for that value, missing value?
A In several of its routines, that is a default.
Q Okay. So let's just assume that Dr. Olsen 11:53AM

0108

intended that to be the case. If you're mistaken about that, though, Dr. Olsen did not substitute the means for missing values when he ran the Sys -- excuse me, the PCA analysis on SysStat, would that change your criticism of this substitution issue? 11:53AM
MR. TODD: Object to form.
A Well, the reason that I'm hesitating is because whether or not Dr. Olsen explicitly did it or accepted the defaults, that was the effect of doing that, and so I find it difficult to 11:53AM distinguish between whether Dr. Olsen did it or he accepted the defaults in the statistical program.
Q Can you run SysStat if we had missing data and it will not substitute the means for the missing data? 11:54AM
A Yes.
Q How do you do that?
A You choose an option.
Q Which option?
A You tell it to not substitute the missing 11:54AM data. You can tell it to do a listwise deletion of the data, which means that if you run into a case that has missing data, you throw it out. You can tell it to substitute the missing value, and I believe that there's a third alternative. 11:54AM

0109
Q Do you know whether or not Dr. Olsen plugged one of those alternatives?
A To be able to run his PCAs and to run the analyses on SW3 and SW15, he had to choose one or the other. 11:54AM
Q Okay. Well, if he chose pairwise deletion, would that affect your criticism on substitution of means for missing values?
A I don't understand what pairwise deletion means. There's -- 11:55AM
Q Okay. Then we'll come back to it.
A Okay.
Q Have you ever reviewed the SysStat manual?
A Yes.
Q You don't recall a decision of pairwise deletion in the manual? 11:55AM
A There's a -- well, the problem is there's not a decision of pairwise deletion. There's a pairwise -- there is a method for computing pairwise correlations. Is that what you're asking me about? 11:55AM
Q Right, and isn't that a method by which SysStat can then use samples with missing data and still perform the correlations?
A Sometimes and sometimes not.
Q Okay. If Dr. Olsen employed that pairwise 11:55AM
0110
deletion method in SysStat, would that affect your opinion on substituting the means for missing values?
A Well, yes and no, and the reason I say that is that in one respect, the calculation of the use of 11:55AM the -- the computation of pairwise correlations, nothing is being deleted. It's a calculation where you are taking whatever observations are available and doing the calculation with just those, which --
Q Right. I understand that. 11:56AM
A Okay. That's why I was objecting to the term deletion because nothing is being deleted.
Q I think that's -- isn't that how SysStat, though, references it in its manual?
A No. I remember listwise deletion, but it 11:56AM would be pairwise correlations.
Q Okay.
A Okay. Oh, anyway, so what I was saying is that the method for pairwise correlations, in some senses if you were to expand the correlations in a 11:56AM Taylor series, the means -- it's sort of like substituting the means. So I have trouble kind of distinguishing between those, but the other problem with use of pairwise co -- that's the no part. It might not change my evaluation of the -- of what I 11:57AM
0111
said about the missing data because if one is an approximation of the other, it would hardly change my conclusions.
The more difficult problem, however, is that if you're using pairwise correlations, which I tried 11:57AM to do, frequently the programs blow up because the correlation matrices are no longer Grammian square. They're are called non-Grammian squared, non, N-O-N, dash, G-R-A-M-M-I-A-N, which means that they no longer have the required characteristics for 11:57AM

symmetry, and it may be that you can't invert the correlation matrices that Dr. Olsen was using, and if you can't invert those matrices, SysStat or SBSS or any of the other packages blow up because you're not able to get the programs -- the routines that they use to convert. So if you did use pairwise correlations, then there's a whole new host of problems. 11:58AM

Q If he did not substitute the means of the values when he did his PC calculations for the scores, PC score calculations -- 11:58AM

A Uh-huh.

Q -- would that affect your opinion on substituting means for -- for missing data?

A Well, it would depend on what he did 11:58AM

0112

differently, but the problem is that something had to go in there because, otherwise, you couldn't get PC scores on all the observations.

Q What if he substituted zero?

A Well, given the way the calculations were done, if he substituted zero, he wouldn't have any results. 11:59AM

Q Just for the missing data for that component?

A You're taking logs. You're taking the log of zero. The log of zero doesn't exist. 11:59AM

Q If you substitute -- instead of putting the mean of that result, if you just substituted zero when you're calculating your PC value, principal component values for your plots, if you just put zero, would that affect your criticism concerning substitution of means? 11:59AM

A Yes.

Q Instead of the mean value --

A Yes, it would.

Q -- you just put zero? 11:59AM

A It would affect my --

Q And how would it affect it?

A Well, that would be even worse than substituting the means because zero isn't a measure of central tendency for anything unless we're 12:00PM

0113

dealing with standardized data, in which case it is the mean.

Q If you are looking for loadings, though, aren't you trying to determine whether or not that particular analyte has a particular impact on that sample, so zero represents a no impact for the sample; correct? 12:00PM

MR. TODD: Object to form.

A Well, it sort of depends on when and where you're substituting to zero. It was -- we were just talking a minute ago, it depends on whether or not it's a standardized value. It depends on whether the analyte is important or not important on that particular principal component. I mean, there are all sort of other factors that you'd have to consider before you decide whether or not zero is important or not. 12:00PM

Q We'll come back to that.

A Okay.

Q Are there anything else -- is there anything else that you would add to this list of key 12:00PM

criticisms you have in Dr. Olsen's report?
A Let's continue. Okay. On Page 26, I have a brief discussion of the non-detects.
Q Uh-huh. 12:01PM

A And I'm going to come back to this later, but the problem with the non-detects is that because non-detect limits differed even for the same analyte because of different test readings. That adds variability to the dataset. That wasn't accounted 12:01PM for.
Q So you suggest here on Page 26 that non-detects should be treated as zero?
A Well, that wouldn't be possible.
Q Well, you say rather than treat this as zero 12:01PM non-detect, Dr. Olsen substitute the midpoint between zero and the detect limit for the chemical; correct?
A That's what I say.
Q So what is your criticism? 12:02PM
A Well, my criticism is that it's not that there is a systematic -- it's not that there is a value substituted for the non-detect; it's that the values vary for even the same analytes. So I give an example, I believe, for aluminum where you've got 12:02PM different non-detect limits, and if there wasn't -- this wouldn't be an issue if the log transforms weren't taken, but once you take the logarithms, those numbers blow up into very large numbers.
Q Okay. What else? 12:02PM

A Okay. We talked before about the difference between the USGS and non-USGS.
Q Right, we already discussed that. Using those -- combining those two datasets?
A Right, without considering whether or not they 12:02PM were the same or different.
Q Okay.
A Okay, and that discussion starts on Page 26.
Q So is this a new or just identifying where you discuss it? 12:03PM
A I'm identifying where I discuss it.
Q Okay. Is there anything else new, new topics?
A Well, we just talked about the non-detect limits, but I do come back to it on Page 75 and talk about sensitivity analyses. 12:03PM
Q Is it paragraph or Page 75?
A I'm sorry. Paragraph 75. I identified concern on Page 33 about the number of principal components and whether or not rotations were used.
Q That's similar to the earlier criticism. You 12:03PM talked about whether or not he used all the principal components or not, or is that a new criticism?
A Actually I don't believe I identified that as criticism. I identified that when you were asking 12:04PM

me about the discussions I had with Dr. Johnson.
Q Thank you. So is that another criticism you have, that Dr. Olsen did not consider all the principal components?
A Yes. Okay. This is less well identified, but 12:04PM in terms of reproducing the data, which I know we

discussed before, I go into details not just about reproducing the data that Dr. Olsen actually analyzed, but I also go into detail starting on Page 35 and running through Page 40. 12:05PM

Q Is this more concerns you have of being able to reproduce the SW3 database?

A No. This has to do with the fact that there's an awful lot of other data that wasn't used and there's not a good explanation about why it wasn't 12:05PM included, and this would include both observations and the variables that were actually analyzed. So -- plus, we go back to the issue that you just alluded to, which is that we're not exactly able to match what Dr. Olsen did. 12:05PM

Q So is the criticism here, there were a lot of samples that were taken that were not used for the PCA analysis?

A Without an explanation as to why they weren't used. 12:05PM

0117

Q Okay.

A Okay. On Page 41, Paragraphs 91, 92, 93, 94, 95 and 96, 97, 98 through 100 and the charts on Page 44 all have to do with the error that Dr. Olsen makes in his calculation of the scores. 12:06PM

Q Okay. What else?

A We just discussed this, but there's further discussion regarding what the analysis would have looked like if a fuller set of variables had been included. 12:07PM

Q And where is that?

A It starts on Page 45 with Paragraph 101, and runs through Page 51 ending with Paragraph 114.

Q Is this when you ran Dr. Olsen's or you ran the PCA using Dr. Olsen's data but including more 12:07PM variables in different samples?

A Yes. I think we tried it several different ways.

Q But your results are on Table 3 on Page 49?

A Right. 12:08PM

Q Okay. Thank you. What else?

A Well, that's all for Dr. Olsen, and then my discussion shifts to Dr. Harwood.

Q Can we wait for Dr. Harwood for now?

A You betcha. 12:08PM

0118

Q Okay, all right.

A Yes.

Q I understand you betcha, too.

A It's okay.

Q That's an Oklahoma colloquialism. 12:08PM

A Yes, sir, and I live in Texas and we have the same one.

Q I want to turn our attention back to this issue of non-detects, which is one of your criticisms; correct, sir? 12:08PM

A Yes.

Q Is it true that non-detects are common when you collect environmental sampling data?

A Yes.

Q Now, I guess -- but you're critical on the way 12:09PM Dr. Olsen treats non-detects in his data; correct?

A Yes. I'm not critical of the fact that he did

something with them because that would be a standard practice in any field, and as I mentioned before, it's common in biostatistics. My concern is the inconsistency of treatment of the non-detects linked to the use of logarithms, and those two things together create a problem. 12:09PM

Q Well, is it your concern that the same analyte may have multiple detection limits; is that the crux of your concern about how Dr. Olsen dealt with non-detects? 12:09PM

A That's a good way to summarize it, yes.

Q What do you understand it -- a lab to mean when they say a substance is non-detect? 12:10PM

A It just means they could not determine that there are any particles in the sample that they looked at, but that doesn't mean that there aren't particles in there; they're just too few to find.

Q Okay. At least by that detection method? 12:10PM

A By that detection -- yes, by that detection method.

Q So you don't believe that a non-detect means there's zero, necessarily zero of that material in that particular sample; correct? 12:10PM

A I just don't know whether it's zero or any other number up to the detection limit, so it could be any number in that range.

Q If you're finding in a set of environmental samples, and I don't know if you understand this or not based on your experience, but let me ask the question. If you are looking at environmental samples and you typically do find a detection for a substance and then you have another sample in the same area, same media where you do have a 12:10PM

non-detect, is it likely that that non-detect probably does not represent zero but there is some portion of that substance in that sample where the non-detect was found?

A Well, that's sort of a general scientific question, and I would say that it would probably be more likely that there's something there that hasn't been detected, but it would also depend a lot on what it was you were trying to detect, how far or near the sample was, its prevalence in the other site and so on. 12:11PM

Q Okay. Now, are you critical because Dr. Olsen substituted zero for non-detects in the dataset; is that your criticism?

A That's not my understanding of what he did. 12:11PM

Q Well, I'm trying to understand what you said on page -- we talked about it just a minute ago. I want to go back to Page 26, Page 57 of your report, Exhibit 1.

A Yes, sir. 12:12PM

Q Would you read the -- yeah, let me restate that. If I said that Dr. Olsen substituted zero, are you critical that Dr. Olsen did not substitute zero for a non-detect when he did his analysis?

A No. 12:12PM

Q Well, would you read the first two sentences under Paragraph 57, please?

A In the data analyzed by Dr. Olsen, he also has a number of values that are non-detects, meaning the measurement method used by the researchers cannot measure any trace measure of a chemical or organic value. Rather than treat this as a zero non-detect, Dr. Olsen substitutes the midpoint between zero and the detect level for a chemical.

Q Okay. Well, when I read that, I assumed that you were critical that Dr. Olsen did not use the word or use the amount of zero for non-detects. Is that not correct? 12:12PM

A I'm not critiquing his use of -- I'm not critiquing his non-use of zero if that makes sense. 12:13PM

Q Okay. Are you critiquing the fact that he used the midpoint?

A No. I'm critiquing the fact that he used different midpoints. The only concern -- I don't have any problem with substituting something for a non-detect. The problem I have is the inconsistent way it's done because of the different detection levels for different tests, and the reason I'm critical of it has nothing to do with providing a value in place of the non-detect. The critique I 12:13PM

0122

have is the fact that if the different tests all have different levels and you're using different non-detect values in the data, then that induces variability.

Q Okay.

A But we're doing principal component -- and it exacerbates the variability because of the logarithms. So the problem is that in use of principal components, when you are trying to explain variability, if you find that a lot of the variability comes from substituting different values and then taking the logarithms as opposed to being more consistent, then some of the results that we're looking at may actually be explaining the variable use of non-detects and not -- have nothing to do with the analyte. 12:14PM

Q So you're not critical of the practice of taking one-half of the detection limit when you run a PCA analysis for a non-detect value?

A The -- I'm not critical of the general method of taking or the tool of taking half the detection limit. What I'm critical of is that the detection limits are jumping back and forth. 12:14PM

Q Okay. I understand that. I just want to make sure we're clear. 12:15PM

0123

A Yes.

Q And isn't it common practice for scientists to take one-half of the detection limit for non-detects when they do PCA analysis on environmental data?

A That's my understanding, and it's true not only for that but in other areas as well. 12:15PM

Q Let me hand you what's been marked as Deposition Exhibit No. 4.

A Thank you.

Q Can you identify that exhibit for the Record, please? 12:15PM

A This is an extract from the book Introduction to Environmental Forensics. That's the cover page,

and then inside it looks like you have a few pages from the chapter written by Dr. Johnson. 12:16PM

Q Okay, and so Dr. Johnson that's on Chapter 7 is the Dr. Johnson you've worked with in this case?

A Well, I don't think I'd characterize it as worked with. Dr. Johnson was working at the same time I was, but I had -- 12:16PM

Q He's another defendants' expert?

A Thank you.

Q And he commented on your report and you commented on his; correct?

A Yes, sir. 12:16PM

0124

Q And do you see Mr. Murphy there on the front?

A I do.

Q Is that the Mr. Murphy that's the --

A No idea.

Q You have no idea, okay. I was going to ask whether or not that's the Mr. Murphy that's also an expert in this case for the defendants. 12:16PM

A I understand that.

MR. TODD: Object to characterization.

A I understand, but I don't know Mr. Murphy. 12:17PM

Q You don't know him, okay. Would you turn to Page 214?

A Yes, sir.

Q Do you see the bracketed area that's been highlighted for you there? 12:17PM

A Yes.

Q See where it says non-detects are indicated in Table 7.3 with the U qualifier; do you see that, sir?

A Yes, uh-huh. 12:17PM

Q And for subsequent numerical analysis, we adopt the common practice of replacing non-detect values with half the detection limit; do you see that, sir?

A Yes. 12:17PM

0125

Q So that's what Dr. Johnson identifies as taking one-half the detection limit of the data that's on Table 7.3; correct?

A Yes, sir.

Q Okay. Let's look to the data on Table 7.3 and let's look under PCB 16. Do you see where the data there has more than one value for non-detects? 12:17PM

A Yes.

Q So Dr. Johnson in this example in his report had multiple non-detect values, and for his PCA analysis, an example used one-half the detection value for this analysis; correct? 12:18PM

A Yes, sir.

Q So isn't your criticism of Dr. Olsen the same thing that Dr. Johnson shows as an example in this report? 12:18PM

A Well, not exactly, and the reason is because in the remainder of the pages that you gave me, I see no indication that a logarithm was taken, and so it's not exactly the same because if you're substituting very tiny values as the non-detects and you are dealing with -- but you're not taking the logarithms, then those values are going to all be relatively close to one another. What I'm objecting 12:18PM

to is this practice when logarithms are taken 12:19PM

because these numbers are no longer close to one another. These number are attenuated by the process of taking the logarithm, and that's what's introduces the variable.

Q Okay. If you take the logarithm of the non-detects shown on Data Table 7A.3 for the first non-detect there, .66? 12:19PM

A Yes, sir.

Q What would that log be for that?

A Log base 10? 12:19PM

Q Yes, sir.

A It would be somewhere -- I mean, I don't do logarithms in my head. Do you just want a range?

Q Yeah.

A Okay. Somewhere between minus one and -- oh, somewhere between zero and minus one, yeah. I believe that's right. 12:19PM

Q Okay. What about the second sample of non-detect, Sample No. 91.3?

A That would be between minus -- let's say it's between minus 1 and minus 2. 12:20PM

Q 1.3 would be one -- minus 1 to minus 2 but .66 would be minus 1 to zero?

A I apologize.

MR. TODD: Would it help to dig up a calculator to make sure the numbers are accurate? 12:20PM

Q Would you prefer to have a calculator to do this?

A Well, that would make life easier if it's okay with you but -- 12:20PM

Q I did not bring mine with me.

A Well, my wife bought me this nifty phone.

MS. HILL: I was going to give him mine.

Q So if you'd like to do the log on there, please do. 12:20PM

A Well, my problem is that I don't know if this calculator takes logs. Does your calculator take logs? My calculator does not take logarithms.

MS. HILL: I think we can find you one.

A Oh, my goodness. Very good. Yes, sir. 12:21PM

MR. TODD: I think we'll give a shout out to the excellent work of the videographer.

A Yes, shout out to the videographer for pointing out that if you turn it, it starts to look like a Hewlett Packard calculator. Pretty cool. 12:21PM

Okay. Thank you. So if I take the log of .66 -- I'm sorry. I'm going to have to run a test here. Okay. So the log of .66 is minus .18.

Q Okay, and point -- 1.3?

A Log of 1 point -- woops. Yeah, 1.3 is .1139. 12:22PM

Q Okay. What about 1.1, which is Sample 17?

A .04.

Q Okay. Would you be concerned if -- with this particular dataset that Dr. Johnson did lab transformations of one-half of these detection limits because it would cause a lot of variability? 12:22PM

A If he did the logs of these?

Q Yes, and then used that in his PCA.

A Well, I would be, although I wouldn't

necessarily look at PCB 16. If I could refer you to PCB 28 --
Q Uh-huh.
A -- which has fewer non-detects, but the values vary more, so they go up to 37 and down to -- I believe the smallest is 1.2. 12:23PM
Q Do you know whether or not in this example Dr. Johnson actually did log transformations of this data?
A I have no idea.
Q And if he did, would that change your view of Dr. Olsen's practice of taking one-half of the non-detect? 12:23PM
A No.
Q So you would think both Dr. Johnson and Dr. Olsen were mistaken in that practice? 12:23PM

0129

A In the specific situation I described before where we have a sizeable range of data and we've taken half the non-detect, the non-detect varies, which it does here on this page, too, and then we take the logarithms, yes, I would criticize that. 12:24PM
Q You would criticize Dr. Johnson that's what he did in this case?
A If that's what he did. I can't help it. It's the same practice.
Q Okay. Let's look at Paragraphs 38 through 41 of your report. 12:24PM
A And I'm happy to do that. I want to add one phrase to the last statement I made. I just want to be sure that we're understanding that I'm talking about the use of the data in the PCA analysis, too. 12:25PM
Q Right. Well, this chapter is -- that we were talking about on Exhibit 4 was PCA analysis; correct?
A Well, the first half is. I read the chapter a while ago, and the first half is PCA analysis. The second half goes on to something else. 12:25PM
Q Well, this part that you read about the common practice of taking one-half of the non-detect values on Page 214, it refers to Table 7.3; correct?
A 7A.3, yes. 12:25PM

0130

Q Okay. So isn't your understanding that this data was an example of data that Dr. Johnson was showing for PCA analysis in this table?
A Well, yes, but I just wanted to be sure that I was clear because the problem is that your pages start on Page 214 talking about principal components analysis, but then this table jumps way over to Page 268, so I don't know at that point whether he's still doing principal component analysis or receptor models. So I just want to be sure that I'm clear that I'm talking doctor -- about the use of this data in a principal components analysis, but I can't state that that's what is happening in this part of the chapter because of the gap. 12:26PM
Q I understand. Let's turn to Page 38 through 41.
A Of my report?
Q Yes, sir.
A I'm going to put this out here if it's okay with you. 12:26PM

Q Actually I just got the five-minute tape. We're into the lunch hour. Why don't we take a break now for lunch before I go to a new topic.
A Thank you.
VIDEOGRAPHER: We are now off the Record. 12:27PM

The time is 12:26 p.m.
(Following a lunch recess at 12:26 p.m., proceedings continued on the Record at 1:32 p.m.)
VIDEOGRAPHER: We are now on the Record. 01:33PM
The time is 3:32 p.m.
COURT REPORTER: 1:32 p.m.?
A I don't think so.
VIDEOGRAPHER: 1:32 p.m.
Q Dr. Cowan, before lunch, we were talking about this issue of non-detects. 01:33PM
A Yes, sir.
Q And it's discussed at least in part on Page 26 of your report. Can we go back to that again? I'm trying to understand your opinion in this area that's contained on Paragraph 57. 01:33PM
A Yes, sir.
Q I'm going to read the first two sentences. I want to start there and then work my way down. Does it not say that in your report, in the data analyzed by Dr. Olsen, he also had a number of values that are non-detects, meaning the measurement method used by the researchers cannot measure any trace measure of a chemical or organic value. Rather than treat this as zero, not detected, Dr. Olsen substitutes 01:34PM

the midpoint between zero and the detect limit for the chemical. Did I read that correctly?
A Yes, sir.
Q Okay. Now, you testified before lunch I believe that you're not criticizing Dr. Olsen by using the midpoint between zero and the detection limit when he ran his PCA, correct, for non-detects? 01:34PM
A I agree, I am not criticizing him for not using zero. Using the midpoint between zero and the lower limit of the detection level is an acceptable procedure. 01:35PM
Q And it's common practice in PCA analysis of environmental data using the midlevel point?
A Well, I don't want to offer an opinion specifically to PCA analysis in environmental data. It's a common procedure used in all of statistics. 01:35PM
Q Okay. What do you mean then, the second sentence, when you say rather than treat this as zero non-detected; what does that phrase add to that portion of your opinion? 01:35PM
A Only that the -- I was offering alternatives because if you weren't taking logarithms, then using zero would be a perfectly acceptable method, too.
Q So if you weren't logarithming, you could put zero in there and that would be an acceptable level 01:35PM

in environmental analysis?
A I think you just created a word. If you weren't taking logarithms.
Q Right. What did I say?
A If you weren't logarithming. 01:36PM

Q Well, I like that word.
A Yeah, it was pretty good actually, but I just wanted to make sure we --
Q Is the answer yes to my question?
A I'm sorry, now I don't remember the question. 01:36PM
Q Okay.
A If you weren't taking logarithms?
Q You would say that then zero would be appropriate as a substitution?
A Zero or the non-detect or the method that Dr. 01:36PM Olsen used.
Q Okay. Let's go on after the first two sentences. However, the detect limits can vary from observation to observation for each chemical; correct? 01:36PM
A Yes.
Q In some samples, we would have smaller non-detects than for others, such as .01 as a lower limit for some observations on aluminum and .001 for other lower limits. Did I read that correctly? 01:36PM

A Yes.
Q Are you suggesting that those are the detect limit observations for aluminum .01, or is that just using an example?
A I'm just using an example. 01:37PM
Q Okay. The next sentence, does it not say this variability in detection limits adds to the variability in the data exacerbated by the use of logarithms. Did I read that correctly?
A Yes. 01:37PM
Q What do you mean by exacerbated?
A Well, the problem is that there's very little variability when you are dealing with numbers like .01 versus .001. They're all -- those are both small values; they're both close to zero. They're 01:37PM not going to impact the overall variability of a measurement where the measurement is coming out in values of 13, 26 or whatever, like in the table that you showed me before. However, when you take the logarithms, what happens is that the .01 becomes 01:38PM minus 2 --
Q Uh-huh.
A -- for the log base 10, and the .001 becomes minus 3.
Q All right.

A At the other end of the extreme, if I'm dealing with a number like 13 --
Q Well, let's just use your examples here. Okay? You got 01 and 02. You'd would have a minus 2 and minus 3 for the log of those two -- 01:38PM
A Right.
Q -- correct?
A Uh-huh.
Q So you're saying that's a greater variability than .01 and .001? 01:38PM
A On the log scale, yes.
Q Okay, but as a matter of fact, the difference between .01 and 001 is a ten-fold difference, is it not?
A Not in a calculation of variability. You're 01:38PM using the ratio, but that's not the way variance is

calculated and particularly not in PCA.
Q But doesn't this normalization process of logarithms tend to take the skewness out of the data? 01:39PM
A Well, you're back to using a term that has a specific meaning in statistics that has nothing to do with this. So could you rephrase your question?
Q Well, doesn't it -- log transformation does normalize the data, does it not? 01:39PM

A You're going to have to define normalize, too, because that has a very specific meaning in statistics.
Q Okay. How would you define it?
A Well, taking the logarithms, essentially for these calculations that are related to variances, tends to take the data from a curval linear space and put it more into a linear space to make it easier to analyze. 01:39PM
Q So using that definition of normalization, isn't that what happened when Dr. Olsen log transformed the data? 01:40PM
A Well, that's one of the things that happens, but the problem is there are other things that are happening at the same time. 01:40PM
Q So you --
A I don't have a problem with taking logarithms either. My problem is specifically the combination of the non-detect limits and the logarithms. I'm not objecting to logarithms independently. I'm not objecting to non-detects. But I am objecting to the fact that the logarithms, combined with the technique that was used for the non-detect, is adding variable to the data, and Dr. Olsen's technique using PCA is designed to capitalize on or 01:40PM

detect variance. So if you are adding variance to the dataset, then you're disguising other things that are going on.
Q Okay. Did Dr. Olsen do any other transformation before he ran the PCA on this data; for example, did he do a Z-transformation of the data? 01:41PM
A Yes, he did. I'm sorry. I'm going to ask if we could repeat the question, though, because the order is important, too. Did you ask me before he did the PCA? 01:41PM
Q Yes.
A At some point before the PCA, yes, he did a Z-transformation.
Q Okay. Wouldn't that then tend to negate any exacerbated variability by doing the Z-transformation? 01:41PM
A No, not at all. It's -- that's just a linear transformation that has no effect at all on the variance, other than the changing the variance for every variable to 1, but that doesn't change the -- it changes the co-variance structure. 01:41PM
Q So you don't think that Dr. Olsen's procedure by Z-transformation reduced the issue you're raising here on Paragraph 57 of your report? 01:41PM

A No.

Q Okay. What about reducing the number of samples with a large number of non-detects; wouldn't that also -- if you didn't use data with log number of non-detects, wouldn't that accommodate the issue you're concerned with here? 01:42PM
A Well, but I'm not complaining about the use of the non-detects. I'm using -- I'm complaining about the use of non-detects with logarithms.
Q Yeah, but what I'm saying, though, is if you didn't have data log non-detects, that would reduce the issue of variability because of the log transform, would it not, because you don't have that many samples that have log transformed data? 01:42PM
A All the samples have log transformed data.
Q Well, non-detects with log transformed data.
A Okay. Now that we've gotten to that, I'm sorry, could we repeat the question?
Q Well, if Dr. Olsen did not use data with a lot of non-detects -- 01:42PM
A Uh-huh.
Q -- wouldn't that reduce the concern of variability you're raising here in Paragraph 57?
A You mean, if he threw away all the data that had non-detects? 01:43PM

0139

Q A lot of non-detects.
A Well, that would create another problem. Well, that would create a couple of problems. One is --
Q Did that deal with the issue you raised in Paragraph 57? 01:43PM
A Would that deal with it? No. It would create a new --
Q Would it reduce the variability, the exacerbated variability you refer to? 01:43PM
A Well, it would eliminate that, but it would increase the variability elsewhere.
Q In fact, when you did the PC analysis, you used data with a substantial amount of non-detects, did you not? 01:43PM
A Yes.
Q Why would you do that?
A Because it wasn't a problem with having non-detects in the data. The problem was the variability that was added by taking the logarithms of the non-detects. So I'll reiterate, I don't have problem with taking -- with accounting for the non-detects. It's a standard procedure in statistics. 01:43PM
Q But if you evaluate correlations of data with a lot of non-detects, is that giving you a valid analysis of correlations? 01:44PM

0140

A If it is -- if you're taking -- well, there are three situations here. There is not taking logarithms, in which case you would have -- the procedure for the non-detects wouldn't really matter that much in terms of computation of variability. Then there's a procedure where instead of not taking logarithms, you are taking logarithms, but the non-detect limit was the same for all the variables. 01:44PM
In that instance, you're consistent in the way you're dealing with the data, and so it adds some

variability but in a consistent fashion across all the variables, and then there's what Dr. Olsen did. What Dr. Olsen did was he used different non-detect levels for different observations and then he took the logarithms, which introduces inconsistency and more variability.

Q Isn't it appropriate in PCA, before you run PCA, that you reduce the skewness of the data, normalize it before you run the PCA? 01:45PM

A Not necessarily. There's all sort of examples where in fact you don't even take the Z transform and you do the PCA on a co-variance matrix instead of the correlation matrix. 01:45PM

Q Well, I assume when you determined that this process that Dr. Olsen did exacerbated the variability, you did some sensitivity analysis to show that?

A I did. 01:45PM

Q And where is that found in your report?

A Well, I give examples later. In terms of the overall sensitivity analysis, I didn't report on that outcome.

Q Did you run -- did you run any sensitivity analysis to show the difference in the results by using log transform of the data as Dr. Olsen did and then do any differently in a method that you recommend here? 01:45PM

A Well, that was kind of a broad statement since you didn't define what it was that I was recommending. So let me tell you what I did. What I did was -- 01:46PM

Q Let me ask you this.

A Excuse me. 01:46PM

Q What are you recommending?

MR. TODD: Counsel, let him ask the question.

Q What are you recommending here?

A Well, you interrupted my response to the earlier question, so I'll answer that one, and then I will tell you what I'm recommending. 01:46PM

Q I'll tell you what. I ask the questions here and you answer them.

A Okay. So then you're going on Record saying that you just keep interrupting me while I'm responding? 01:46PM

Q No. You indicated to me you did not understand my question. That's a fair concern. So let me try to restate it.

A Great. Thank you. 01:46PM

Q Now, when you were complaining about exacerbation --

A Uh-huh.

Q -- did you do any analysis comparing Dr. Olsen's method, which you've criticized -- 01:46PM

A Yes.

Q -- did you run a different method to show that it really did exacerbate the variability of the data? 01:47PM

A Yes.

Q What did you do?

A I ran the data without -- with the same

non-detects but without the logarithms.
Q Okay, and where are those results? 01:47PM

A They're in my work papers, which I provided you.
Q You didn't report that here in the paper?
A No.
Q And what did they show? 01:47PM
A They showed that you get different results if you run it without the logarithms.
Q Were they substantially different results?
A In some instances, yes, you get substantially different results in two respects. You get 01:47PM different results because different variables cluster together, which completely undercuts the concept of having a signature, and then in addition, the ordering of the principal components changed because you're explaining different variance 01:47PM components.
Q Okay, but you don't have those results in your report?
A Not in this report. It's in the materials I you were provided earlier. 01:48PM
Q Could you find them this evening and bring them to the deposition tomorrow?
A I didn't bring them with me. I provided all of my information, but I didn't bring all my files with me. 01:48PM

Q How would you identify that in your files?
A You can find a number of files all labeled according to which dataset was used, SW3 or SW15, and then whether or not logarithms were taken or not taken, and in some cases in those files you would 01:48PM find whether or not there are rotations.
Q Okay. If -- if the data was not normalized in this case by Dr. Olsen, wouldn't coliforms totally dominate the analysis?
A You're going to have to define which 01:48PM normalization that you're talking about.
Q Log transformation and then the Z-transformation before PCA we've been talking about here I thought for the last few minutes.
A Okay. Well, yeah, but those are both 01:48PM considered normalizations. I wasn't sure which one you were referring to.
Q Both were performed before he ran his PCA analysis; correct?
A Yes. 01:49PM
Q Okay. So I'm saying if he didn't perform those transformations --
A Right.
Q -- wouldn't coliforms totally dominate the analysis? 01:49PM

A For taking the Z score or not taking the Z score, I believe if you didn't take the Z score, you would get the result that you talked about with coliforms dominating the analysis.
Q Let me hand you what's been marked as Exhibit 01:49PM 5 to your deposition.
A Thank you.
Q Can you identify that, sir, for the Record?

A It says expert report of Brian Murphy and then has several lines that essentially describe the case. There's the case number and it's signed by Brian Murphy. That's the first page, and the second page talks about multimedia principal component analysis. 01:50PM

Q Have you ever reviewed Dr. Murphy's report? 01:50PM

A No.

Q Okay. Well, on Page 30 Dr. Murphy discusses a PCA run of this data, and it mentions that Dr. Murphy, in the third paragraph down, used one-half of the detection limit for non-detect results. Do you see that, sir? I should have highlighted it for you on your -- 01:50PM

A You did, thank you, and I do see that.

Q It doesn't state whether or not Dr. Murphy log transformed the data, but if he did so before he ran his PCA analysis, would you be critical of the way Dr. Murphy ran his PCA? 01:50PM

A Well, I can't answer your question for a variety of reasons, but one of the reasons is that this also doesn't state whether or not he had multiple non-detect limits. 01:51PM

Q Well, he's using the same dataset Dr. Olsen used.

A That doesn't mean --

Q Assuming that's the case -- 01:51PM

MS. COLLINS: Object to form.

Q -- and he used multiple detection limits --

MS. COLLINS: Object to form.

Q -- I want you to assume that for me.

A Well, okay. Well, you're -- as long as we're clear that I don't have any idea what Dr. Murphy did, and so if you're posing a hypothetical that he used the same non-detect limits as Dr. Olsen; is that your question? 01:51PM

Q Yes. Treated in the same manner and also did the same transformations before he did his PCA, would you be critical of the way Dr. Murphy did his PCA? 01:51PM

MR. TODD: Object to form.

A Well, keeping -- there are several other hypotheticals that are going on here. One is that I haven't seen the rest of what he's done, and I haven't reviewed any of the work by Dr. Ol -- or by I'm sorry, Dr. Murphy. So I'm merely speculating based on one paragraph that you've shown me. 01:52PM

Q I've given you the hypothetical.

A Okay, but giving me that hypothetical, that hypothetical is sort of the same hypothetical you gave me when you showed Dr. Johnson's data, and then we weren't sure whether or not he was taking the logs there, too, but at that time I indicated I'd be critical of Dr. Johnson if he was taking the logarithms of the data with multiple different -- multiple and different non-detect limits, and so I would have to say the same thing for Dr. Murphy. 01:52PM

Q Let me hand you what is marked Cowan Exhibit No. 6.

MR. PAGE: Did you get the marked number there, marked No. 6 on there?

COURT REPORTER: Yes.
MR. PAGE: Thank you.

Q Have you ever seen that document before, sir?
A No.
Q Would you read for the Record what the title page is?

A Numerical Ecology, Second English Edition by two professors, who oddly enough are named, Legendre.
Q Maybe they are brothers.
A I was referring to the fact that there was a famous mathematician several hundred years ago named Legendre. 01:53PM
Q Okay. Would you turn to the second page of the exhibit? Well, I'll note for the Record that this was -- this exhibit was introduced in the preliminary injunction hearing by the defendants that cross examined Dr. Olsen. 01:54PM
Do you see the statement there, misuses of principal components?
A Yes, sir. 01:54PM
Q Would you read the second full paragraph into the Record, please?
A Principal component analysis was originally defined for data with multinormal distributions, Section 4.4, so that its optimal use, Cassie and Michael 1968, calls for normalization of the data, Subsection 1.5.6. 01:54PM
Q Okay. Do you agree or disagree with that statement you just read?
A Well, there are a number of different things in this statement, so -- 01:54PM
Q Let me be more specific.
A Okay.
Q Do you agree or disagree that the optimal use for principal component analysis calls for normalization of the data? 01:54PM
A Well, the problem that I'm having is with the predicate, which says that principal component analysis was originally defined for data with multinormal distributions, and the optimal use that it's referring to there has to do with once you go from data that is multinormally distributed to the principal components so that you get to a summary matrix of scores that would be distributed as a Wishart distribution, then I would agree that that would call for a normalization of the data, but there's no claim here that -- that this data is multinormal, number one, and there's no tests performed. So we don't care about what the probability of distribution is. 01:55PM
Q Okay. The next sentence, would you read that, please? 01:55PM
A Deviations from normality do not necessarily by the analysis, however, Ibanez 1971, and then there's a period. So this isn't a complete sentence. 01:56PM
Q Okay. Would you read the next sentence, please?
A It is only important to make sure -- or maybe

I just read the sentence wrong. I apologize. I need to do over. Deviations from normality do not necessarily bias the analysis, however. It is only important to make sure that the descriptors' distributions are reasonably unskewed. 01:56PM

Q Okay. Do you agree with those two sentences, that statement? 01:56PM

A I do, but the skewness that is being referred to has to do with distribution of the data, the probability of distribution of the data.

Q And not how you unskewed data through log transformations? 01:56PM

A Well, see, nobody uses the term unskewed data when you are talking about doing transformations because of the confusion with the term skewness with respect to probability distributions. It's a measure of the third moment of the distribution. 01:56PM

Q So you're saying that this does not refer to the activities that Dr. Olsen did by log transforming the data to reduce skewness of the data before he ran PCA? 01:57PM

A Okay. Well, my problem is that you keep using skewness in a way that's completely inconsistent with what this paragraph is talking about. This paragraph is talking about measurements of the third moment of a probability distribution. You're talking about whether or not the data has extreme values. They're two completely different concepts. 01:57PM

Q Okay. So then do you agree or disagree with the two sentences you just read?

A I agree with the statements that are read here. It just has nothing to do what you're asking me before. 01:57PM

Q Would you read the next two sentences, please?

A Typically an analyses conducted with strongly skewed distributions, the first few principal components only separate a few objects with extreme values from the remaining objects, instead of displaying the main axes of variation of all objects in the study. 01:57PM

Q And so are -- is it your testimony, sir, that that statement also does not relate to what has gone on in this case, what Dr. Olsen did in this case? 01:58PM

A Again, since everything else in this paragraph has to do with skewness of the third moment of probability distribution and since Dr. Olsen didn't look at the probability distributions, made no claims about the normality or non-normality of the data and didn't do any probabilistic testing, I don't see how it has anything to do with anything other than what this is referring to, which is a probability statement. 01:58PM

Q Does this have anything to do with the effect of coliform or bacteria data on the dataset that we're evaluating in this case?

A No, because this is talking about a probability distribution and the transformation of the normal distribution to the Wishart. 01:58PM

Q Would you turn to Paragraphs 38 through 41 of your report. Do you have that, sir?

A Yes, sir. 01:59PM

Q What is the concern that you're voicing on these portions of your report?
A Okay. Well, starting with -- I'm sorry, I'm going back one page to find out what the table is. Okay. The table at the top of Page 38 -- 02:00PM
Q Top of Page 38? I'm sorry?
A Yes, sir.
MR. TODD: Did you say Page 38 or Paragraph 38 when you asked him to --
Q I thought we were looking at Paragraph 38. 02:00PM

Did you turn --
A Oh, I'm sorry. Yeah. I took Pages 38 to 41.
Q I'm sorry. I meant -- if I misspoke, I apologize. I'm talking about Paragraphs 38 through 41. 02:00PM
A I may have misheard, and I suggest we don't go back and have them read it. That way we can both be right.
Q Just take a moment to read through that and I'll ask you something. 02:00PM
A Yes, sir.
Q Could you summarize, please, what the concern is you're raising in this portion of your report?
A Okay. We touched on this this morning. The way Dr. Olsen treated the data that he had available 02:02PM was -- for most of the analytes, you only have one observation, so it is whatever it is, one measurement. So if you are looking at copper, you got one number, but for the bacteria there were multiple samples taken and analyzed to measure the 02:02PM bacterial levels, whatever -- whichever the four measurements, and they varied by quite a bit.
Q Is this for all the bacteria samples or just for a few samples?
A Well, it varies from sample to sample to 02:03PM

sample. So sometimes there are four observations for bacteria. For -- let's stick with just one, so, you know, fecal coliform.
Q I just wanted to understand how often this occurred. How many times did you find samples with 02:03PM multiple observations of the same analyte or organic material?
A As I understand it, it's most of the time, but I never counted that up.
Q Most of the time? 02:03PM
A So -- because what happens is that frequently there are multiple observations, but it's not always four, sometimes it's three, sometimes it's two, sometimes it's one, sometimes -- I think we got up to five but -- 02:03PM
Q Can you tell me what kind of samples were involved in these cases?
A I don't understand your question.
Q Were they duplicate samples or split samples?
A My understanding is that one -- this is -- my 02:03PM understanding is that a water sample was taken, and in the water sample, multiple extractions were taken from that water sample to measure the coliform levels.
Q Okay. So that was a -- can we call that a 02:04PM

split sample then for terminology?
A Why don't we call it multiple subsamples because split sample means something else in sampling theory.
Q What's the difference? 02:04PM
A Well, a split sample means there's a formal division that you do because, for example, you want to compare the left side to the right side. So if I have --
Q Okay. I understand. That's not what we're doing in environmental -- 02:04PM
A No, we're not doing that.
Q And this is an environmental case. You understand that, do you not?
A That's why I asked if we could use the word subset sample. 02:04PM
Q Well, in environmental cases this would be either called a split sample or a duplicate sample.
MS. HILL: Object to the form.
MR. TODD: Object to form. 02:04PM
A Well, I understand that, but since my background is sampling theory, I answered your question regarding what most sampling theorists would call it.
Q Okay, and in this particular case, though, 02:04PM

what we're talking about here is taking one sample of the same media at the same time and place and then divide it into multiple --
A Yes, sir.
Q -- pieces to analyze; correct? 02:05PM
A Yes, sir.
Q Okay. That's -- we'll call that a duplicate sample?
A I'm happy with that.
Q Okay. Well, is it -- the appropriate way to deal with a duplicate sample is to average the results from the same analyte for that sample? 02:05PM
A Well, you need to define appropriate because if you are talking about is it an appropriate procedure, can I do it? Yes. Is it an appropriate procedure so that I then plug it into a PCA when I'm looking for variability? No. 02:05PM
Q Well, if you did not combine and average such samples, wouldn't you give too much weight to that point in space and time versus other sample observations? 02:05PM
A That's the way the sample was drawn. I don't get to choose.
Q Well, do you understand that oftentimes you'll check analytical methods by labs by doing split 02:06PM

samples and having the same lab analyze essentially those two samples at the same time to check the analytical integrity of the lab?
MR. TODD: Object to form.
A Well, naturally I understand that. That's not what the issue is. The issue is that if you have all of those observations and then you take an average, you're making believe that all the information that you just got from having the lab do that work is summarized in a single statistic, and it's not. 02:06PM

Q But you do agree that -- so you're saying that that would be inappropriate to average those together -- let me ask the question again. So do you believe it would or would not give it too much weight if you did not average those duplicate samples --

MR. TODD: Object to form.

Q -- in the analysis?

A I do not know how to answer that question. You're going to have to ask me that question in a different way. 02:07PM

Q Well, if you did not average the duplicate sample results --

A Uh-huh.

Q -- in the PCA analysis, wouldn't you give too much weight to that point in space and time versus other sample observations?

A Well, first of all, many of the sample observations had multiple observations on the bacteria. Secondly, averaging is just one way to deal with that. There are many other ways to conduct the PCA that wouldn't require averaging at all. 02:07PM

Q Okay, but is averaging an appropriate way to deal with that?

A No. I just got through saying it's not appropriate because you're doing a principal component analysis that's supposed to be looking for variability. 02:07PM

Q So how would you avoid the -- placing too much weight on that sample and that point in space and time if you did not average?

A Bootstrapping.

Q And what is that? 02:07PM

A Bootstrapping is a procedure that's used in estimation to deal with situations like this where you take -- bootstrapping is an unfortunate name but that's the name. It is a method where you draw samples out of your sample, but you do this 02:08PM

repeatedly, like 5,000 times, 10,000 times, and analyze each of those separately, and then look at the distribution of the results that you got as a measure of the outcomes.

Q Do you know that other environmental professionals, such as Dr. Johnson, treat duplicate samples the same way Dr. Olsen does? 02:08PM

A I don't know that. Could we take a 30-second break?

Q We can take five minutes if you like. 02:08PM

A Thank you.

VIDEOGRAPHER: We are now off the Record. The time is 2:08 p.m.

(Following a short recess at 2:08 p.m., proceedings continued on the Record at 2:13 p.m.) 02:13PM

VIDEOGRAPHER: We are now on the Record. The time is 2:13 p.m.

Q I want you to look again back to Exhibit No. 5, which is a partial part of the -- Dr. Murphy's expert report in this case. 02:13PM

A Yes, sir.

Q The same area we read before, does it not say

multiple results for the same sample for the same analyte were averaged?
A Yes. 02:14PM

Q So if Dr. Murphy had employed this process of averaging duplicate samples, the same analyte, would you be critical of Dr. Murphy's analysis also?
A Yes, for the same reason.
Q Okay. I'll hand you No. 7, sir. 02:14PM
A Could I add one brief phrase to my last statement, which is assuming that the samples were different sizes.
Q Assuming the samples were different sizes?
A Yes. He says multiple samples. If, you know, 02:15PM in one case it's four observations on bacteria and in another case it's three, and that's what Dr. Murphy did, then, yes, I had the same concerns.
Q Okay. Thank you.
A If it's four, four, four, four, four, 02:15PM it's harder to --
Q It's my understanding that Dr. Murphy was using the same data that Dr. Olsen was.
A And that may be, but I can't tell that from there. 02:15PM
Q I understand that.
A I wanted to be clear that I was talking about the same situation.
Q Can you tell us what Exhibit No. 7 is, sir?
A It's an article from the Canadian Journal of I 02:15PM

guess Fisheries and Aquatic Sciences in 2000 entitled Defining the Sources of Airborne Polychlorinated Biphenyls: Evidence For the Influence of Microbially Dechlorinated Congeners From River Sediment, and it's asked as a question. 02:16PM
Q Okay, and is Dr. Glenn Johnson, defendants' expert in this case, a co-author of this report?
A He is a co-author of this report. I'm assuming it's in the same Glenn Johnson.
Q Is this a case where -- well, let me ask you 02:16PM this: I want you to turn to Page 90 of the report, please.
A 90?
Q See where it says statistical treatment? Do you see about two-thirds of the way down, the top 02:16PM left paragraph where it says, duplicate samples were averaged before statistical manipulation and use in tables and figures. Do you see that statement, sir?
A Yes, sir.
Q If Dr. Johnson employed that method, would you 02:17PM also be critical of Dr. Johnson for averaging duplicate samples?
A Well, I haven't read the rest of the article, so I don't know what he's using it for. Keep in mind that I was criticizing its use in principal 02:17PM

components, which is what I assume Dr. Murphy was doing in the previous thing that you showed me, but in this case, since I haven't read the article, I don't know that he's doing principal components analysis. 02:17PM
Q So you'd only be critical if he did the averaging of the duplicate samples before principal

component analysis but not other types of environmental investigations of source?
A Well, it depends on the type. I mean, that's such a broad statement it's almost impossible to respond to. It depends on the type of analysis being done. 02:17PM
Q Dr. Cowan, let's make sure, how often did you actually see four analysis of the same parameter or variable? 02:18PM
A You mean four observations --
Q Yes.
A -- that were taken for --
Q The same sample, in the same -- 02:18PM
A You know, I don't really remember at this point. I mean, we're talking about the initial go-through of the data that occurred months ago.
Q Was it more than ten times?
A I don't remember. 02:18PM

0163

Q Let me hand you what's been marked as Exhibit No. 8.
A I don't think you need to give me two copies.
Q That's great. Thank you.
A Sure. 02:19PM
Q Are you familiar with John Davis' text, Statistics and Data Analysis in Geology?
A No.
Q Would you know whether or not this is the most well-known and leading environmental geochemical data, statistical text? 02:19PM
A I'd have no idea.
Q Would you turn to Page 35, please, and the first full paragraph, would you read the first sentence, please? 02:19PM
A In geochemical analysis, it is common practice to make multiple determinations or replicates of a single sample. The most nearly correct analytical value is taken to be the mean of the determinations.
Q Do you agree or disagree with that statement by Dr. Davis? 02:20PM
A I would say that my -- this looks like it's a very introductory text. So I would say that if you were trying to represent the measure of central tendency and that was your purpose in doing this, 02:20PM

0164

I'd say I would agree with the statement. If you're planning to do something else with it, then I would disagree.
Q So if Dr. Olsen was trying to represent the measure of central tendency for that particular sample in that analyte with multiple observations, it would have been appropriate to take the mean -- 02:20PM
MR. TODD: Object to form.
Q -- of those samples?
A Well, I understand your question, but that's not my understanding of what Dr. Olsen wanted to do. 02:20PM
Q Would you answer my question, please?
A If Dr. Olsen were attempting to represent a measure of central tendency in the sample --
Q Yes. 02:21PM
A -- then that would be fine.
Q Thank you. I want you to turn, please, sir, to Page 29 of your report, and I want to focus on

Paragraph 66.
A Yes, sir. 02:22PM
Q Would you read the first sentence?
A Dr. Olsen doesn't explain why he takes logarithms. He simply does so.
Q Would you explain what you mean by that?
A Sure. Somewhere else in this document I 02:22PM

0165

give -- oh, on the previous page and the top of this page, I give three reasons for taking logarithms, and any of those three reasons relates to the purpose of or the -- how taking logarithms fits in with the statistical analysis. I don't understand 02:22PM necessarily why Dr. Olsen takes logarithms. There's not a sufficient explanation from a statistical perspective as to why he does so.
Q Okay. Did you carefully read Dr. Olsen's report? 02:23PM
A I did.
MR. PAGE: No. 9?
COURT REPORTER: Yes.
A Thank you.
Q What I've handed you is the cover page and 02:23PM then a few of the pages from Section 6 of Dr. Olsen's report beginning at Page 641. Would you, please, review the Pages 641 and 642 and tell me whether Dr. Olsen on those pages explained why he log transformed his data. 02:23PM
A You just want me to look at 641 and 642?
Q Yeah, the top of Page 641 --
A Yes, sir.
Q -- where we kind of bracketed it there for you. 02:25PM

0166

A Okay.
Q Doesn't that explain why Dr. Olsen log transformed his data?
A Well, it explains why he transformed the data. It doesn't explain why he log transformed the data, 02:25PM first of all. Secondly, although I understand that Dr. Olsen believes it's desirable to have distributions that are near normally shaped -- but let's go through the two reasons. First of all, there's no hypothesis testing of any sort done in 02:25PM Dr. Olsen's report, so I don't understand why it's important to have data that's near normally shaped --
Q Are you stating --
A -- for any probability distribution.
Q Excuse me. I just want to be sure. Are you stating that Dr. Olsen doesn't state any hypothesis to be tested in his report?
A No, I didn't say that. I said he didn't do a formal hypothesis test where you use probability 02:26PM distribution.
Q Well, in the first page there on Page 641, does it not say in conjunction with the descriptive statistics listed above, probability plots or P plots are generated in order to examine the 02:26PM

0167

distributional shape of the data for each variable?
A That's what it says, but you don't need probability plots to do that, and probability plots

don't tell you anything more than the other types of plots that you might run, like just regular, you know, scatter plots. 02:26PM

Q Did I misunderstand you a few minutes ago that you testified that Dr. Olsen did not review probability plots in his analysis?

A That's not at all what I said. I said he didn't conduct any formal probability-based tests. Reviewing a probability plot by the eye is not the same as conducting a formal algebraic test. 02:26PM

Q So you're of the view that it's better to look at the numerical results rather than look at the eye results on the probability plot to determine the data analysis? 02:27PM

A You know, I didn't say that either. That was something that you just made up whole cloth out of something else I said. So that was a complete mischaracterization of everything I just said. 02:27PM

Q Well, I wasn't trying to mischaracterize what you're saying, Dr. Cowan. I'm trying to understand what you said.

A Okay. Well --

0168

Q You seem to dismiss the analysis of looking at probability plots, and I'm saying isn't that as valid of an analysis as you are suggesting?

A It is -- well, it would be, except for the fact that a probability plot doesn't have anything to do with normality. You could get the probability plots -- in fact, you would get the probability plots from the data before you ever took a transformation. So I don't understand what the transformation has to do with looking at probability plots. 02:27PM 02:28PM

Q So you're suggesting now that Dr. Olsen did not suggest -- in these paragraphs that he did not explain why he log transformed his data; is that correct? 02:28PM

A Well, let me point out, in the previous sentence that we skipped over, possible transformations available in ED analyzer, so right away we're relying on an Excel spreadsheet tool.

Q Can you direct me where you're looking, please? 02:28PM

A Oh, I'm sorry. I'm at the top of Page 641, first full paragraph, the third sentence.

Q Okay.

A Okay. He mentions that there are four 02:28PM

0169

possible transformations available in ED analyzer. Let's be clear that there are actually hundreds of transformations that are available. He's only describing four that happen to be available in his Excel spreadsheet. However, these transformations are natural logarithms, base 10 logarithm square and square root. There -- going from there to saying and that's why I took the base 10 log transformation, there's no explanation of why choosing -- 02:28PM 02:29PM

Q Well, he says right here, this step is important for PCA for two reasons, does he not?

A Excuse me. You're reading it in backwards order. What he says is transformations are

important, and then this step of taking a transformation is important in the PCA for two reasons. He doesn't say anything about base 10 log transforms. He could have easily -- as easily taken the square root transformation and avoided several of these problems. 02:29PM

Q Would you read below there where it starts to say in practice; would you read that out loud, please?

A In practice for most of the PCA runs, data were base 10 log transformed for all variables, although there were exceptions to obtain near normal distributions for most of the parameters and to minimize the effect of highly variable concentrations in units of measure. 02:29PM

Q Doesn't that explain why Dr. Olsen log transformed the data? 02:30PM

A No. All it does is -- it does not because he just got through saying that there were other transformations available. The square root transformation would have avoided many of the problems that he has with the non-detects, and he said that was a good transformation, too. So why take a base 10 log transform as opposed to a square or square root --

Q Well, he said --

A -- or a tangent.

Q But your point was on Paragraph 66 of your report, he doesn't explain why he takes logarithms; he simply does so?

A Right. 02:30PM

Q And what you just read is he gave two reasons why he did log transforms; is that not correct?

A No, it's not correct. He gave two reasons for taking transformations, and then he says he took the log transforms. He doesn't explain out of the four 02:30PM

that he offered or the hundreds of others that are available why he chose logarithms versus any other type of transformation.

Q Well, but he does explain why he used log transforms, does he not, for two reasons? 02:30PM

A No. You've asked me that three time, and each time I've said no.

Q So to obtain near normal distributions for most of the parameters and to minimize the effect of highly variable concentrations in units of 02:31PM

measures --

A You can -- -

Q -- those aren't reasons?

A You can keep reading it back to me out of order as you are, and I keep telling you he gave 02:31PM that sentence related to all of the transformations and then he says that he took log transformations. So reading it out of order only gives me the information in this paragraph that I'm looking at out of order. It doesn't explain anything. 02:31PM

Q Do you also claim that Dr. Olsen did not run any statistical tests on his data?

A I'm sorry. If I did, you're going to have to show me where it is.

Q Paragraph 77 of your report. 02:31PM

0172
A Paragraph 77?
Q Excuse me. Let me check this.
A This talks about rotations.
Q Oh, I'm sorry. I said 77. I meant paragraph -- we're back on 62. I think it's 62, last sentence of Paragraph 62. I apologize. 02:32PM
A That's okay.
Q Would you read that, please?
A As Dr. Olsen didn't conduct any statistical tests, this can't be the reason. 02:32PM
Q Are you claiming that Dr. Olsen did not conduct any statistical analysis on this data?
A Okay. Well, you just changed it from tests to analysis. He obviously conducted a statistical analysis. 02:33PM
Q Okay. So you're claiming he didn't conduct any statistical tests?
A That's what the sentence says.
Q Okay. What do you mean by that?
A Well, typically when you have, excuse me, a set of hypotheses such as the ones that Dr. Olsen laid out, you then follow that up by conducting statistical tests to determine whether you accept or reject the hypothesis. I'm sure Dr. Olsen conducted other types of tests elsewhere, but there aren't any 02:33PM
0173
tests to formally state whether his hypotheses are true or not true.
Q So is it your position, Dr. Cowan, that a statistical test, such as a t-test, would be better than a visual evaluation of a probability plot? 02:34PM
A For what purpose?
Q To test his hypothesis.
A Well, it depends on the hypothesis, but with regard to the hypotheses that were the primary hypotheses in this analysis, you couldn't examine a probability plot. You would have to conduct a formal t-test to determine whether or not, given the size of the samples that were used, you could accept or reject the hypothesis. You'd never be able to tell that by looking at a probability plot. 02:34PM
Q Do you know whether or not Dr. Olsen did a preliminary data analysis to justify the use of log transformations?
A I don't know.
Q Would that be important to your analysis? 02:34PM
A Well, it might be or it might not be, and it would depend on whether or not Dr. Olsen looked at the other transformations that he also suggested and then said based on these factors, I choose this one over these others over here. 02:35PM
0174
Q Let me hand you what's been marked as Cowan Deposition Exhibit No. 10. This is Appendix E from Dr. Olsen's report. Have you reviewed that prior to this day, sir?
A I need to go through the document first, please. 02:36PM
Q Thank you. Have you reviewed this before today, sir?
A I did when I first got the report, yes.
Q Okay. Would you agree that the analysis shown 02:37PM

in Appendix E does explain why Dr. Olsen did log transformation on his data?
A No.
Q Why not?
A For the very reason I cited before you handed me the document. It makes no comparison of the log transformed to any of the transforms that Dr. Olsen himself cited. There's no comparison of these log transforms to the square root, to the square, to other types of transformations that might have been conducted. It's just a set of approximately 20 charts that give me what the probability plot is without any discussion about why this might be better and/or worse than the square root transformation and, furthermore, some of the data in 02:37PM 02:37PM 02:38PM

0175

here indicates to me the square root transformation probably would have been better.
Q What data is that?
A Well, if you look towards the end, for example -- I'm sorry, the last chart in the document you gave me -- I'm sorry, it's not the last one. It is -- here it is, total dissolved solids, which is fourth from the back. That's not even remotely close to a probability plot for a log distribution. It's heavily influenced by a number of outliers, and so the transformation just takes it further from a normal distribution. So why not consider a different distribution that would achieve the two goals that Dr. Olsen stated in the document you gave me earlier. 02:38PM 02:38PM 02:39PM
Q Well, look at the probability plots for bacteria.
MR. TODD: Counsel, where are those in here?
MR. PAGE: Well, I didn't number the pages, so we'll have to pick through here. It's E. coli, Enterococcus, then fecal coliform. 02:39PM
A You want to look at coliforms?
Q Fecal coliforms, yeah. Doesn't that show that the Log10 transformation was appropriate, that 02:39PM

0176

probability plot?
A Well, all it says is it is one of several probability or it is one of several transformations that may have worked, but if you go back a couple of pages to calcium, it shows that it didn't work. So this isn't an analysis. This is a set of charts. An analysis provides actual practical, rational thought to the process in terms of conducting analysis. Printing pages that are just charts isn't an analysis; it's printing pages. 02:39PM 02:40PM
Q Are you suggesting that's what Dr. Olsen did, he just printed pages?
A That's what I'm suggesting. Dr. Olsen printed pages and didn't give any thought to whether or not it's a normal distribution or a normal transformation, a square root transformation or any other kind of transformation that would meet the two goals that he stated he had. 02:40PM
Q But would you agree or disagree that transforming -- log transforming sampling data for statistical analysis of environmental data is a 02:40PM

common practice?
A Well, sure, but that doesn't mean it's the correct practice in this case. It's just a common practice. 02:40PM

0177

MR. PAGE: This is number --
COURT REPORTER: 11.
Q Let me show you what's been marked as Exhibit No. 11. Are you familiar with this particular text, Statistical Methods For Environmental Pollution Monitoring? 02:41PM
A No, sir.
Q You've never seen this before?
A No.
Q You wouldn't know whether this is the leading text on environmental statistics or not? 02:41PM
A No, sir.
Q Would you turn to Page 164, please, Chapter 13. What's the title of Chapter 13?
A Characterizing Lognormal Calculations. 02:42PM
Q Would you read the first sentence, please?
A Lognormal distribution is the most commonly used probability density model for environmental contaminant data.
Q Do you have any basis to agree or disagree with that statement? 02:42PM
A You do realize that this is talking about a probability distribution that has nothing to do with this case, the lognormal?
Q Could you please answer my question, Dr. 02:42PM

0178

Cowan?
A Okay. I apologize. What was the question?
Q Did you -- do you have any basis to agree or disagree with that statement?
A Well, based on everything else I've seen so far from the other documents you gave me, I'm not sure I'd agree. 02:42PM
Q So you disagree with the statement?
A No. Once again, you're mischaracterizing what I said. What I said was I'm not sure I'd agree. I didn't say I disagreed. I'm saying that relative to all the other documents you've shown me, this is the first time the lognormal distribution has been brought up as a probability distribution, and the other documents you gave me discussed the normal distribution and Wishart distribution. 02:42PM 02:43PM
Q So do you -- let me ask it this way then: Do you agree with the statement -- the first sentence on paragraph -- the first paragraph on Page 164?
A I don't have any way to disagree or agree. 02:43PM
Q Let me hand you what's been marked as Exhibit 12. This is the same portion of the same textbook we referred to earlier, right, that was by Dr. Murphy?
A I assume so. 02:44PM

0179

Q Okay. Would you turn to page -- the second page of Paragraph 136 -- Page 136, the bottom paragraph.
A Yes.
Q Would you please read the bottom paragraph up to the point where the reference is to Ott? 02:45PM

A Although most -- I'm sorry, you're talking about this last paragraph?
Q Yes, sir.
A Thank you. Although most statistical tests are based on the assumption that the underlying distribution is normal, most environmental data appear to have frequency distributions that are lognormal. Two advantages of the lognormal distribution in describing environmental data are that it always gives positive values. There are no negative concentrations, and it can account for a small fraction of higher values, hotspot contamination in the right side or tail of the curve. 02:45PM
Q Do you agree with those statements?
A I do.
Q Doesn't that statement support the use by Dr. Olsen of log transformation of his data?
A No. You have completely confused taking a logarithm with a probability distribution that happens to have the unfortunate name lognormal. Taking a logarithmic transformation of data does not suddenly make it lognormal. It starts out as lognormal and you analyze it that way. Dr. Olsen's data was lognormal when he started. He didn't have to take a log transformation to get it into the lognormal distribution. You're talking about two concepts that are so totally far afield that it just demonstrates that you have no idea what a probability distribution is relative to a transformation of data. 02:46PM
Q When I take a logarithm on the data, is that not the first step for doing a lognormal transformation? 02:46PM
A No. That's taking a logarithmic transformation. A lognormal distribution, which is what is being described here, is a probability distribution that has characteristics related to the normal distribution but has nothing to do with logarithmic transformations. It just is lognormal. This is also the most commonly used frequency distribution in financial analysis for the exact same reasons, but nobody is taking logarithms of the data. They start out by assuming that it's 02:47PM
lognormal because of the characteristics that are described here, and it's used to estimate extreme risks, several papers I've published on.
Q Isn't that lognormal distribution a transformation done in order to reduce the skewness of the data? 02:47PM
A You obviously are just not even remotely listening to what I'm saying. Lognormal here is referring to a type of probability distribution that's characterized by a specific function that has nothing to do with logarithms. Okay? Dr. Olsen is taking a logarithm transformation of the data, which transforms it to get it to look like it's normally distributed, which is a completely different process, a completely different problem and comes out of two completely different areas of mathematics. 02:47PM

MR. TODD: Could we take a quick break?
MR. PAGE: Sure.
VIDEOGRAPHER: We are now off the Record. 02:48PM
The time is 2:47 p.m.
(Following a short recess at 2:47 p.m., proceedings continued on the Record at 2:55 p.m.)
VIDEOGRAPHER: We are now on the Record.
The time is 2:55 p.m. 02:56PM

0182

Q Dr. Cowan, during our break, we -- I handed you what's been marked as Exhibit 13. It's another copy -- parts of John Davis' text?
A Yes, sir.
Q Would you turn to Page 97. It's the second page of the exhibit. 02:56PM
A Yes, sir.
Q There it also refers to a lognormal law just above the figure, does it not?
A That's part of a section that you've highlighted, yes. 02:56PM
Q Okay. Could you read the highlighted section for the Record, please?
A Yes, sir. The pattern --
Q Yes, sir, right there. 02:57PM
A Yes, sir. I'm just making sure that I understand what this sentence is referring to since it starts out the pattern. The pattern comprised of a minimum value lower limit, a low background -- low level background containing the bulk of the 02:57PM
observations, a tail of decreasing numbers of observations having higher concentrations and a few anomalies whose concentrations may exceed the background by orders of magnitude is so ubiquitous that it has been called the lognormal law of 02:57PM

0183

geochemistry. Such distributions can be transformed to a more tractable shape simply by taking the logarithms of the concentration values as shown in Figure 2-41B.
Q Taking that last sentence there, such distributions can be transformed to a more tractable shape by simply taking the logarithms of concentration values as shown on Figure 2-41, could you explain that statement to me, please? 02:57PM
A Sure. Any distribution that looks like what he characterized in here with a minimum value lower limit, which just means that there is a lower limit that exists that's not minus infinity, some background with the bulk of the observations and the tail presumably going to the right with higher 02:58PM
concentrations and then a few anomalies, that describes maybe 20 or 30 different probability distributions, the F-distribution, the Chi-squared distribution and the t-distribution that we discussed before, all those distributions look 02:58PM
exactly like what is there. Okay? He's saying that you can take logarithm of the data to put it into what he calls a more tractable shape without explaining what a tractable shape is.
Q Okay. Does he not demonstrate what a more 02:59PM

0184

tractable shape is below in the Figure 2-41?
A Yeah. I was referring to what tractable

means.
Q Okay. In effect, isn't what Dr. Olsen did is take a logarithm of this data the same as suggested here on Page 97 of Davis' report? 02:59PM
A Possibly.
Q Okay. Was there -- would there be any benefit to transform or take a logarithm of these environmental data before you do any environmental analysis? 02:59PM
A Well, yeah, but Dr. Olsen also pointed out that there was several other transformations that he could have taken, too.
Q Okay, but there is some benefit to do the logarithmic transformation? 02:59PM
A Well, sure. I gave you an example earlier.
Q Thank you, sir. Could you turn to the next page, please?
A Uh-huh. 03:00PM
Q Could you help me understand what Figure 2-42 is?
A Sure.
Q Let me ask this: Are these probability plots?
A Not in the same sense. They are probability 03:00PM

0185

plots, but they're completely different than the ones that we were discussing before.
Q They're different than the ones on Appendix C for Dr. Olsen's --
A Yes. 03:00PM
Q -- analysis? How are they different?
A Well, this shows the cumulative distribution, excuse me, of data, so it is actually a different probability measure that looks at the cumulation of the observations as opposed to the distribution of the observations. 03:00PM
Q And the difference is these are cumulative distributions as opposed to single observations?
A Yes, sir, and they're both cumulative distributions. The only difference between the left side chart and the right side chart is that on the right side he has presented the axis as a logarithmic axis. 03:01PM
Q Although the right side chart, isn't that saying, gee, the left side chart has just been log transformed or logarithmic transformed? 03:01PM
A Well, that's the distinction. It's not that there's a logarithmic transform. It's that he's actually changed the scale of presentation.
Q Okay. So it's changed the logarithms on the 03:01PM

0186

right; correct?
A Yes, but I don't -- but I'm trying to distinguish between changing the data versus changing the scale of presentation. So he hasn't taken logarithms. What he's done is he's changed the scale under which the data is presented. 03:01PM
Q Isn't that what you do when you go from an arithmetic scale to a logarithmic scale, just do a logarithm on their arithmetic data?
A Sometimes but not always. I mean, it sort of depends on how you are presenting the data. 03:02PM
Q Is that what happened here in Dr. Davis' text on Page 98?

A Well, I'm having a little trouble determining that because I can't actually find points that translate from the left to the right based on what you are describing.

Q Okay. Would you turn to Page 43, Paragraph 99 of your report, sir?

A Sure. 03:03PM

Q Would you read the first sentence, please?

A I'm sorry, of Paragraph 99?

Q Yes, sir.

A Finally, since the SysStat values are still on logarithmic scale, the proper interpretation of the 03:03PM

0187

values would be on a real-world scale. This is easily done --

Q Could you just stop there for a moment?

A Sure.

Q Could you explain that first sentence you just 03:03PM read; what do you mean by proper interpretation of the values? I guess the logarithmic values from SysStat would be on a real-world scale; is that what you are suggesting?

A I have to -- oh. What I'm saying is that the 03:04PM values that came out of the analysis at the end were still on the logarithmic scale, but it's not obvious that they're on the logarithmic scale, and so it would be -- since we started with real-world values, we transformed them by taking the logarithms. Then 03:04PM we standardizes the values. Then we did the PCA. Then we computed the scores, excuse me, and unstandardized, and now we're still in the log stage. All I'm saying is you started with real world. You went through logs, and then you went 03:04PM through the Z transform. You did the computations. You calculated the scores. You untransformed them to get them back onto the -- to undo one of the two transformations, and now they're -- but they're still in log values, so why not take them out of the 03:05PM

0188

log values back to the real world?

Q What do you mean by -- well, are you suggesting that you would interpret PC scores on a real-world scale?

A Well, that's what is commonly done. 03:05PM

Q Aren't PC scores dimensionless? I'm having difficulty understanding what you mean by real world. When I hear the words real world, I think of something has dimensions.

A Uh-huh. 03:05PM

Q So are PC scores -- do they have dimensions; aren't they dimensionless?

A No. They're new dimensions. They're -- each one is an axis that is computed through the cloud of points that you've got, and the new axis is what you 03:05PM are determining is your PC scores. So in that respect, it is a dimension.

Q But aren't they really comparative on a relative basis as opposed to an absolute basis?

A Well, yeah, but that doesn't have anything to 03:06PM do with the logs versus the non-log values.

Q So I'm not -- I'm sorry, I just don't understand what you mean by converting them and comparing them on a real-world scale. What would

you have done? 03:06PM

A Well, let's ignore the standardization for a minute. Okay? What we started with was measures on 26 different variables. So I've got copper; I've got iron; I've got all those things. Okay? Then I do some other stuff, including calculating the PC scores. Oh, I'm sorry. I've got the original measures on copper, iron and so on. Okay? So each one is a measure of how much is there. 03:06PM

Q All right.

A Okay. Now I take the log of that, for whatever reason. Okay. Then I do my principal components analysis, and when I'm done with the principal components analysis, because I did the principal components analysis on the log values, at the end, whether I standardized or not, I'm still on the log scale. So in other words, I haven't gone the one final step from log back to measurement of how much there is. 03:07PM

Q Is the PC score intended to be a measurement of how much there is? 03:07PM

A Well, it depends on what you're trying to do. In this case since I thought that the whole point of this was the presence or absence of chicken litter, yes.

Q What would you have measured? 03:07PM

A I'm sorry?

Q What does the PC score measure? Are you saying it measures chicken litter?

A Well, wait a minute. I'm not the one doing the analysis. So all I'm saying is that as an example, the PC score in this case is measuring how much there is of these different analytes. Okay? So we started with how much copper and iron there was. Okay? Then we go off and we do this analysis, and we find, for example, a factor that says that iron, copper and two other metals are commonly found together. That's my interpretation of -- 03:07PM 03:08PM

Q And the strength of the score, the PC score would be the strength of that correlation, correct; it wouldn't be how much concentrations there are? 03:08PM

MR. TODD: Object to form.

Q Isn't that true?

A Well, not exactly. What it does is it uses the correlation, but it indicates that if there's a lot of one, there's a lot of another. 03:08PM

Q Where there's a little bit of one, there's a little bit of another?

A That's -- yes. That would be at the other end of the dimension.

Q Right. So I'm trying to understand how you would take your PC score and translate it into concentrations. 03:08PM

A Oh, I wasn't claiming you should translate it into concentrations. I was just saying you should take it out of logarithms. 03:09PM

Q And what would you put it into then; what would it represent then?

A Well, if -- well, that's part of the problem that I've got. I'm not sure what it would represent

exactly, but I'm saying if you started with real measures and real concentrations and you did a transformation to get to the logs and then did this analysis, and then you come out of it and you still have the logs, I don't understand why you're not going the final step and saying I got into this by taking the logs, I did this transformation for a specific reason, and now I'm going to go through this analysis, and at the end I'm still on the logarithmic scale because all of my inputs were on a logarithmic scale, but the real data I started with was not on that scale, it was on a completely different scale, and so I'd like to go back to that one. 03:09PM

Q So what would you do to go back to that one as you are referencing here in Paragraph 99? 03:09PM

0192

A I would take the -- well, actually I did the calculation for you here.

Q Okay. What was that? Right here in this paragraph?

A No. There's -- I have to find the equation, so if you'll excuse me just a second. No. Way more complicated than that. Here it is. If you would kindly go to Pages 29 and 30. 03:10PM

Q Okay.

A Okay. In Paragraph 67 I present the generic equation for principal component. 03:10PM

Q Yes, sir.

A I just say principal component is written as blah, blah, blah, blah, blah. The Vs in this equation are actually the logged values of the concentrations. 03:11PM

Q Okay, sir.

A Okay. Well, if that's the case and -- so I repeat that at the top of Page 30, okay, and what I have then is a principal component that is the sum of the values C Sub A times the log V Sub A. So if I've got variable one, I've taken the log and I multiply it by weight, and that's a principal component that Dr. Olsen has computed. 03:11PM

Q Okay. So where do you show us how to 03:11PM

0193

transform the data as you mention in Paragraph 99 --

A Okay.

Q -- the real-world scale transformation?

A Okay. If you go to Paragraph 69 --

Q Uh-huh.

A -- I have a principal component that is -- just by doing regular old algebra, I have the log of this product. Okay. Well, if the principal component is the log of that product, then it's still in the log space, so I can take the principal component scores and take 10 raised to the principal component score, and that would be the reverse transformation. 03:12PM

Q And so that becomes a real-world score?

A That would be the score that's not in the log space. So I'm calling it real world, but I'm just trying to say that it's not the logarithmic space. 03:12PM

Q You take it out of logarithmic --

A Yes, sir.

Q -- scale? 03:12PM

A Yes, sir.
Q Okay. In this particular example, you showed me on Page 29 and 30, do you see any Z-transformation of the data?
A No, because this is after the 03:13PM

0194

Z-transformation. So Z-transformation has already been undone.
Q Okay. So you're assuming the Z has already been undone?
A Well, actually that's what Dr. Olsen did, so, 03:13PM yes.
Q Do you know of anyone who does this transformation that you're suggesting when they do a PCA analysis?
A No. 03:13PM
Q Do you know whether or not Dr. Johnson, when he does PCA analysis, transforms his data back into this real-world arena?
A That was a good description. I don't know. I didn't discuss this with him. 03:13PM
Q You're not aware of anyone else who has thought that way?
A Well, you're asking me about environmental.
Q Yeah, environmental.
A And since we determined earlier since this is 03:14PM my first environmental case, no.
Q Can we please turn to Page 19, Paragraph 44 of your report?
A 19, yes, sir.
Q Let me catch up with you here. Excuse me. 03:14PM

0195

I've got some of my things out of order now. Okay. I want you to, please, direct your attention to the last two sentences of Paragraph 44, and would you read those to the Record, please?
A This means only 47 percent, less than half, of 03:15PM the observations have real data actually observed in the field. This means that more than half of Dr. Olsen's observations have data that Dr. Olsen substituted rather than real data.
Q Okay. I think earlier you might have 03:16PM explained this. Would you explain what you mean that only 47 percent, less than half, of the observations have real data actually observed in the field?
A Sure. 03:16PM
Q I mean, isn't it true that all of the observations that Dr. Olsen used had real data in them; is that correct?
MR. TODD: Object to form.
A Well, all of the observations had some real 03:16PM data, and more than half of them also had plugged-in data.
Q So wasn't it an exaggeration on your part to say that less than half of the observations had real data actually observed from the field; isn't that an 03:16PM

0196

exaggeration?
A No, sir. It's exactly -- that's the computation that comes exactly out of that and that I then use later in this report.
Q Well, all of the observations of Dr. Olsen 03:17PM

have real data in them, do they not?
A They all have some real data.
Q Okay, and some of them have as many as six of 26 parameters missing; correct?
A I believe that's correct, yes. 03:17PM
Q Okay. For the SW3 dataset, that's the surface water dataset; is that correct?
A That's one of the surface water datasets.
Q Okay. That's the one that Dr. Olsen ran his principal component analysis for his opinion in this 03:17PM case, on surface water?
A Well, actually I believe he ran analyses on both SW3 and SW15.
Q Okay, and that was for missing data or datasets without missing data; correct? 03:17PM
A Yes, sir.
Q But the one he based his opinion on was SW3; correct?
A I believe that's correct, yes.
Q The total missing values, do you know what the 03:18PM

0197

total missing values are from the SW3 dataset, that is, if you looked at all the potential variables or parameters that are within those observations, how many total missing values were there?
A I knew at one point. I don't recall, and 03:18PM you're asking me I believe about the cells.
Q I might be able to help you.
A Thank you. Is it in my report?
Q Yeah, I think it is.
A Okay. 03:18PM
Q I think it's on Page 37. Let's look at Page 37 of your report, and I think if you look at Paragraphs 84 and 86, it may refresh your recollection.
A I'm sorry, 84 and 86? 03:19PM
Q Yes, sir, or 84, 85 and 86.
A Okay.
Q 86 is just simply I think --
A Two charts.
Q -- your charts, yes, sir. And my question is, 03:19PM while you're looking at that, Dr. Cowan --
A Okay.
Q -- is actually how many missing values are out of the total values potential values from these observations? 03:19PM

0198

A Well, in reality or in Dr. Olsen's SW3?
Q In Dr. Olsen's PCA analysis SW3 dataset.
A 915.
Q Okay, and that's out of how many potential observations or observations in the dataset? 03:19PM
A I believe I counted 14,898.
Q Okay. So is that approximately 6 percent? I'll refer you to Paragraph 86 of your report.
A Yes, sir.
Q So of Dr. Olsen's analysis for PCA SW3 surface 03:20PM water, there was a total of 6 percent missing values from the observations he used in his run; correct?
A Missing cells, yes.
Q Missing parameters or variables?
A Well, no, because the whole variable was 03:20PM there. I'm thinking of this as a big spreadsheet,

okay? Down the rows of the spreadsheet, which is actually what Dr. Olsen had, down the rows of the spreadsheet you had the observations. So an observation is --

Q A sample?

A Yes, a sample. The columns are the 26 columns that give the different --

Q The different analytes?

A Right. So there are -- so that's why I was referring to cells because it's the cells in that table, and there are 14,898 of those, of which 915 were empty in Dr. Olsen's SW3. 03:21PM

Q So in environmental terminology, there would be 915 missing analysis from the 14,898 potential analysis for those observations? 03:21PM

A As I understand the environmental terminology that we've been discussing today, yes, I would agree with that.

Q And do you recall, sir, that at least 20 of 26 parameters -- that all of the observations in Dr. Olsen's PCA for SW3 had to have at least 20 of the 26 parameters that were to be used? 03:21PM

A I don't remember that exactly. I'm sure that I said something about that in my report, too. 03:22PM

Q Does that sound familiar to you, sir?

A Actually, no.

Q Does not?

A Well, the 26 variables sounds familiar. The fact that 20 of them were complete does not. 03:22PM

Q Okay.

A It just seems high given the fact that half the observations had some missing data.

Q Okay. Maybe an observation might have one item missing? 03:22PM

A Well, that's true, but the problem is that if 53 percent of the observations had some missing data, Dr. Olsen had another rule, which was that he wouldn't accept a variable if it didn't have a minimum amount of data in it. So I don't think it's possible to get back to 20, although -- I mean, it's possible. It just doesn't seem likely. 03:23PM

Q Do you know that Dr. Olsen does discuss that in his report on --

A Yes, and actually I thought I had, too. I apologize. I just don't remember where. 03:23PM

Q Did you review Dr. Olsen's sensitivity analysis where he evaluated SW15 and SW 16 to SW3, that is, he ran the PCA with the missing data and without -- and with observations without missing data; did you evaluate that? 03:23PM

A Well, I remember reading it, yes.

Q Okay. Do you recall forming any conclusions as to whether or not the fact that there was 915 missing data points have a substantial impact on the PCA analysis? 03:24PM

A It had some effect, although I don't believe that, for example, it reordered the presentation of the principal components that were found.

Q Did it reorder the coefficients? 03:24PM

A I simply don't remember.

Q Would that be important to evaluate whether or not this is a critical issue in the PCA?
A Well, reordering the coefficients?
Q Yes. 03:25PM
A It depends on how significantly they were reordered. If they jumped from principal component to principal component, that would be a significant issue.
Q You don't recall whether you did that analysis or not? 03:25PM
A Well, I remember that Dr. Olsen did. Is that your question?
Q Yeah.
A Okay.
Q And did you feel that his analysis was sufficient to justify the use of the observations with 20 of 26 -- no more than 20 of 26 -- no more than 6 of 26 missing observations?
A Well, I understood that -- I understood and actually agreed with that standard. The problem is that I don't believe that Dr. Olsen adhered to that standard. 03:25PM
Q You don't believe that he adhered to the standard that there were no more than six missing cells for each observation? 03:25PM
A Yes. In a different part of my report I note that there are -- when we went through the data, the full dataset on the axis data file that you and I discussed earlier, we found more observations than that. So Dr. Olsen moved some observations that would have qualified under that standard. 03:26PM
Q So you're saying that there are some samples that had 20 of 26 observations or data points and that were not used in Dr. Olsen's SW3? 03:26PM
A Well, I don't want to adhere to the 20 because we didn't agree to that number before, but if -- hypothetically if that's the number, then, yes, there are more samples than that.
Q Do you know when you found those additional samples that may have qualified for SW3, whether those samples had rejected data in them or not? 03:26PM
A I know what you mean by rejected data. I went over that with Dr. Reeves, and I believe that they do not because I remember -- I remember having that discussion with Dr. Reeves. It's like, well, how did they get in there and, you know, aren't they part of this other problem, and he said no. 03:27PM
Q So he -- so you're saying that you found some samples that did not have rejected data that would have qualified for Dr. Olsen's PCA analysis under SW3? 03:27PM
A Well, I think I'm saying actually even a little bit more than that, which is that we weren't sure why Dr. Olsen focused on the 26 variables that he had since there were originally over 300, there were 315, and we found that there were 56 variables that had -- that met those criteria, so there's -- 03:28PM
Q Met what criteria? Excuse me.
A The criteria you were describing before about having no more than six empty cells. 03:28PM
Q Let me hand you what's been marked as

Deposition Exhibit 14. Can you identify that document, sir?
A I'm sorry, I've confused myself on my papers. Yes, sir. 03:28PM
Q Okay. Would you -- have you seen this document before?
A Well, it's another part of Dr. Olsen's report, so, yes. 03:29PM
Q Did you review it?
A Yes, sir.
Q Would you look to this third page of this document; do you see where it says SW3?
A Yes, sir. 03:29PM

0204

Q That is in fact the surface water PCA analysis that Dr. Olsen ran; correct?
A Yes, sir.
Q How many total samples does it show?
A 573. 03:30PM
Q And that's what you found also; correct?
A As to what Dr. Olsen did.
Q Yes.
A Yes.
Q Okay. Do you see there that it shows the number of parameters over parameter threshold, the 26 of 20? 03:30PM
A Yes, 26 and 20.
Q Does that refresh your recollection as to the criteria that Dr. Olsen used? 03:30PM
A Yes. Thank you.
Q So you agree with me now that the -- it had to be 20 of 26 parameters in order to fit into this?
A Yes, sir. It's also repeated in SW3 where he has more observations, 607. 03:30PM
Q Do you know what it means by without SPLP?
A No, sir.
Q Excuse me?
A No, sir, but they both say without SPLP.
Q And would you turn to -- oh, several pages back where we get to SW15. 03:30PM

0205

A Sure.
Q Was that the run where you used just the samples where he had all 26, 26 observations?
A Yes, sir. That's the same result we found. 03:31PM
Q Okay, and is it your testimony that the results, the PCA results did not substantially change between those two runs?
A Well, actually what I said was that the principal components extracted came out in the same order. 03:31PM
Q Okay. Thank you.
A Okay.
Q Excuse me for a minute. Let's look at Page 23, please. 03:32PM
A Yes, sir.
Q Is it your opinion that Dr. Olsen substituted mean values for missing data in that sample?
A At some point, yes.
Q What do you mean by at some point? 03:33PM
A Well, in our discussion this morning, you indicated that there was some confusion about whether Dr. Olsen did it, whether he accepted the

defaults within SysStat, whether he used paired comparisons, but in any case, to get the final set of PC scores that he calculated outside of SysStat, he had to plug in the mean values to get those scores. 03:33PM

Q For missing values?

A Yes, for the missing values. 03:33PM

Q Okay. Is it also your opinion that this practice was a source of error in Dr. Olsen's analysis?

A Yes.

Q Would you look at the first two sentences on Paragraph 54 -- excuse me, the first two sentences on Paragraph 45. 03:34PM

A Yes, sir.

Q Would you read those, please?

A Sure. This is the third key problem in Dr. Olsen's analysis. Dr. Olsen has plugged in so many missing values that a very significant part of the dataset is made up by Dr. Olsen. 03:34PM

Q So is this, again, restating that you believe he substituted missing values with mean values? 03:34PM

A Yes.

Q Okay. Would you read the next sentence?

A While he analyzes both the dataset with no records with missing data and the second dataset with substituted data, he fails to admit that he's 03:35PM plugged in values that skew the correlational structure.

Q Read the next sentence.

A Dr. Olsen substitutes the mean for a missing value: If the aluminum is missing, he substitutes 03:35PM the mean for aluminum from the other sampling sites where aluminum was recorded.

Q Would you read Footnote 6, please?

A There is no direct statement in the CDM report that states missing values are replaced with means, 03:35PM but replacing missing values with means is the only way to reproduce the results from Dr. Olsen's analysis.

Q So what's your -- so Dr. Olsen does not ever state that he replaced missing values with means, 03:35PM but it's your opinion it's the only way you can run SysStat is to do that?

MR. TODD: Object to form.

A No, sir, that's not what I said.

Q Well, what do you mean by Footnote 6 that says 03:35PM there is no direct statement in CDM values that state missing values were replaced with means, but replacing missing values with means is the only way to reproduce the results in Dr. Olsen's analysis?

A Well, what we discussed this morning was that 03:36PM there were multiple options within SysStat, the default being to plug in missing values if you -- I'm sorry, the mean value for missing, but the -- another option is the one that you referred to this morning as paired correlations. The fact is, is 03:36PM that Dr. Olsen had to do something about the missing data because, otherwise, the SysStat wouldn't run. It would just stop and say I don't know what to do.

So it establishes a default, which is what Dr. Olsen could have chosen, or if he cared to override the default, then he could choose a different option. He could have also structured the Excel data files because, remember, we started with --

Q Doctor, let me stop just a second, Dr. Cowan. I'm just focusing on substituting missing values with averages. Okay? 03:37PM

A Yeah.

Q Are you suggesting that Dr. Olsen, before he ran SysStat, substituted missing values with average values? 03:37PM

A I'm saying that I don't remember whether he substituted those or whether the SysStat substituted those, but that's --

Q Okay. Well, you say in your report -- you accuse him of substituting missing values with the means; correct? 03:37PM

A Well, he had to at the end.

Q Well, no. You said this -- were you referring to Page 45 -- you're saying on Paragraph 45 that this was at the end? 03:37PM

A No. I'm saying that he had to at some point.

Q Okay. Well, so you -- it's not true, is it, that Dr. Olsen -- you don't know whether Dr. Olsen substituted mean values for missing values before he ran his SysStat analysis, do you? 03:37PM

A Okay. I believe that we're parsing words here because the problem is that to get the results that Dr. Olsen has, you have to plug in missing values, and that's for the PCA analysis.

Q Okay. That's what you believe, that's what you believe, you have to plug in the missing values with mean values; that's what your opinion is right now; is that correct? 03:38PM

A Yes.

Q Okay. So even though you don't know for sure, you believe that's what Dr. Olsen must have done; he substituted the mean values for missing values before he ran his PCA analysis; correct? 03:38PM

A That's what I believe because it was the only way that we could replicate Dr. Olsen's work. Now, if Dr. Olsen substituted something else, then he needs a justification as to what it is that he substituted.

Q Well, do you realize that you can run SysStat with pairwise deletion in circumstances where there's missing data so you do not have to substitute mean values for missing data in order to run the SysStat? 03:38PM

MR. TODD: Object to form.

A And do you remember the discussion -- 03:39PM

Q Could you answer my question yes or no?

A Yes.

Q So you can. And, in fact, let me show you exhibit -- let me show you what's been marked as Exhibit 15. Can you identify this document? 03:39PM

A It is documentation describing the operations of SysStat, and this particular book is one of, I think, ten that come with SysStat. This particular one is the base statistics book.

Q Okay. Would you look at the second page of the exhibit under missing data? 03:40PM
A Yes, sir.
Q Would you read the first three sentences there, please?
A Ordinarily factor analysis and other multivariate procedures delete all cases having missing values on any variables selected for analysis. This is listwise deletion. For data with many missing values, you may end up with too few missing complete cases for analysis. Select pairwise deletion if you want co-variances or correlations computed separately for each pair of variables selected for analysis. 03:40PM
Q So if you have missing data, you can select pairwise deletion under SysStat and run your PCA analysis without substituting mean data for missing data; is that correct? 03:40PM
A Well, you may or may not be able to.
Q Okay. Well, did you try to do this in your analysis to see if it would run? 03:41PM
A I did.
Q And were you able to -- did you select pairwise deletion when you ran the analysis?
A I did.
Q And what happened? 03:41PM
A It blew up because it's not Grammian square, and there was a sufficient amount of missing data that caused the programs to blow up.
Q Now, did you run it with Dr. Olsen's 573 data points pairwise deletion; did you try to replicate 03:41PM
--
A Yeah. Wasn't that the SW15?
Q SW3.
A I'm sorry, 3. I got them backwards. Yes.
Q Okay, and that's in your -- did you show that result in your report? 03:41PM
A No. It's in my work papers.
Q And where is it in the work papers?
A It's one of the many Excel files that were included in the work papers given to you. 03:41PM
Q So it's your testimony, sir, that you selected pairwise deletion for SW15 and it did not work?
A This was such a great example --
Q For SW3?
A Thank you. This was such a great example that I actually called everybody into my office and held it up as an example of why you wouldn't want to do this because it caused SysStat and then SPSS both to blow up. 03:42PM
Q What do you mean by blow up? 03:42PM
A It means that the program --
Q I mean, the machine blew up?
A It was very messy.
Q You mean it really did blow up on you?
A No, sir. 03:42PM
Q You tend to have a tendency for hyperbole. I'm trying to understand the precision of your word, sir.
A Okay. Well, blow up is a jargon term that

refers to computer systems that quit running and crash. So it wasn't hyperbole. It was use of a jargon term describing what happens if you try to do something that's nonsensical and the computer quits working. 03:42PM

Q So you're suggesting that what the SysStat manual recommends here for running datasets with missing data is nonsensical?

A It is if their program -- they're the ones who suggested it and then the program didn't run. It doesn't make sense to me. 03:43PM

Q So SysStat's manual is nonsensical?

MR. TODD: Object to form.

A You know --

Q Yes or no.

A You can't answer yes or no to that question. 03:43PM

Q Well, what I'm trying to understand is, is that it appears to me from this Exhibit 15 that SysStat recommends selecting pairwise deletion if you want co-variances or correlations computed separately for each variable selected for the 03:43PM

0214

analysis. Are you suggesting that that -- when you do that under SysStat, it will blow up?

A No. I'm saying that it will blow up sometime. You don't -- but it blew up in my case. Would you have suggested that if that happened to Dr. Olsen, 03:43PM that he turn in a report that says, sorry, can't do this because pairwise deletion blew up?

Q Do you know whether or not Dr. Olsen successfully used pairwise deletion when he did his SW3? 03:44PM

A I don't know whether he did or not, but I know when I tried it, it blew up.

Q How many times did you try to run it?

A Twice. I also had occasions when it worked.

Q Oh, so you had some occasions when it would work and some occasions when it wouldn't work? 03:44PM

A Yes.

Q So that's your complete answer now is that it blew up one time and other times it worked?

A No. I said more than once. 03:44PM

Q So how many times did you run pairwise deletion?

A I think four times on different datasets I had.

Q And how many times did it work? 03:44PM

0215

A Twice.

Q How many times did it not work?

A You want me to do the computation for you?

Q Yes, sir. You're the mathematician.

A Thank you. That would be two. 03:44PM

Q Thank you. So is it still your opinion that Dr. Olsen must substitute mean values for missing data before he runs his PCA analysis in SysStat?

A Well --

Q Yes or no, sir. 03:45PM

A Well, I can't answer that yes or no.

Q Let me hand you what's been marked as Exhibit 16.

A Thank you.

Q And that's another portion of Dr. Murphy's 03:46PM

report. Again, I know you haven't seen that report; correct?
A Yes, sir.
Q Would you look at one, two, three, four paragraphs down from the top and read that into the Record, please? 03:46PM
A Sure. Using the dataset described above as input for the PCA run, the output from SysStat reproduced the coefficients reported by Dr. Olsen. This is true only when based on a correlation matrix 03:46PM

with pairwise deletion.
Q So it would appear that Dr. Murphy was able to run SysStat without it blowing up when he ran the SW3 dataset using pairwise deletion; correct?
A Yes. 03:47PM
Q Let's look again back at Dr. Murphy's report there on Page 25.
A Sure.
MR. TODD: Which exhibit are we on now?
MR. PAGE: Exhibit 16, still on 16. 03:49PM
A It would be Exhibit 16.
MR. TODD: Okay. Great.
Q Dr. Murphy states that he was able to reproduce the coefficients reported by Dr. Olsen in his report for SW3; is that correct? 03:49PM
MS. COLLINS: Object to form.
Q Does he say that or not?
A I'm sorry, because I didn't hear everything you said.
Q Yeah. Isn't it true that in the area that you just read a minute ago on Exhibit 16 -- 03:49PM
A Sure.
Q -- that Dr. Murphy states in his report that he was able to reproduce the coefficients reported by Dr. Olsen for SW3? 03:50PM

A Yes.
Q I hand you what's been marked as Exhibit 17. Do these show -- does this exhibit show the coefficient results that Dr. Murphy is referring to in Exhibit 16? 03:50PM
MS. COLLINS: Object to form.
A Kind of. It gives a graphical representation of the values, but it doesn't give the actual numbers.
Q Just to make sure I understand, when we look at the coefficients on these bar charts, does it not provide the numbers for the coefficients there? 03:51PM
A I'm sorry. I didn't see that page.
Q I'm sorry. It's a lot of pages.
A Oh. No. That's what I was saying. I can't tell you whether that's a .06 -- 03:51PM
Q Oh, because it's hard to find it on the --
A Yes. That's all I meant.
MR. TODD: Could we identify which page we're looking at? 03:52PM
MR. PAGE: Yeah. It's page -- it's actually Figure 6.11-11.
MR. TODD: Thank you.
Q But if Dr. Murphy is correct and he was able to reproduce Dr. Olsen's results for SW3 by using 03:52PM

pairwise deletion, then he would have obtained the results that are shown on Exhibit 17?
MS. COLLINS: Object to form.
MR. TODD: Object to form.
A Well, let me point out that I also got those results. I was able to reproduce Dr. Olsen's results and never said I wasn't able to. 03:52PM
Q Oh, so you were able to reproduce the results that are shown on Exhibit 17?
A Well, actually I did it more exactly because the -- that -- I can't tell what the coefficients are. I had his coefficients exactly. 03:52PM
Q Were you able to reproduce his results exactly using pairwise deletion?
A No. I substituted the means. 03:52PM
Q And you got the same results?
A I did.
Q Where could we find that file in your considered materials?
A That would be -- if you remember earlier today, we discussed the SW3 files logged, unlogged, rotated, not rotated. I mean, that's how they are labeled. They're in my work papers. 03:53PM
Q Do you remember what directory it's in?
A I wasn't aware that we gave it to you in directories. 03:53PM
Q So you just found just one massive work paper file?
A You know, I hate to tell you this --
Q I'm trying to understand -- 03:53PM
A I hate to tell you this --
Q -- where you are telling me to look.
A -- but I believe that that may have been the case, but there's -- if it's in a directory, there are multiple directories and only some of them have analytical results. 03:53PM
Q Okay. Now I want to talk about computation of the PC scores after the PCA was done.
A Oh, yeah.
Q You've done your PCA. 03:53PM
A Okay.
Q Now you want to look at your actual scores, compute your scores --
A Sure.
Q -- of your different observations?
A As one of the outputs from the principal component operation?
Q Right.
A Sure.
Q So now you want to compute your PC scores. 03:54PM
A Okay.
Q When Dr. Olsen computed the PC scores, and this is after the analysis was performed, isn't it true that he substituted the mean of the log transformed in standardized data a value of zero for missing values, which resulted in a contribution of zero in the PC score calculations? 03:54PM
A Before removing the standardization, that would be true.
Q Therefore, the missing data contributed nothing to the scores; correct? 03:54PM

A No, sir, that's not true.
Q It's your testimony --
A I'm sorry.
Q -- they contributed to the -- to the contribution -- contributed to the scores for the PC observations? 03:54PM
MR. TODD: Object to form.
A Well, yeah, because you standardized. So using the Z scores, what you did was you plugged in the middle value, but then the next step in his operation was to unstandardize, which moved that value, whatever the zero was, back out to whatever the original scale was before standardization. So it's hard to say it contributed nothing, given that 03:55PM

0221

the very next step was to move that zero point to something else.
Q Well, my question said isn't it true that he added a value of zero for the missing values?
A Well, no, I'm sorry, that wasn't your question. That was part of the first question, but that wasn't the full question. 03:55PM
Q Okay. So back to the first question to make sure I understand, if you just add a zero rather than the mean for the missing values, doesn't that indicate that there was a contribution of zero for the PC score calculations? 03:55PM
A Okay. Well, I'm sorry. Your -- your -- the way you described it is a little difficult to respond to. It wasn't that he put in a zero instead of the mean. He put in the zero because it was the mean. He standardized. So by standardizing with a Z score, the mean of any transformed variable that's on a Z score is zero, okay, but then when you unstandardize it, that moves it back out to whatever place it holds on the real number line, you know, 35. 03:56PM
Q And what is your criticism of that? We don't have a value, so the mean of the value is zero, and that's what's being attributed to that score? 03:56PM

0222

MR. TODD: Object to form.
A My objection to that is the same objection I had in my report, which is that the mean for a value -- let's say you are missing aluminum but that you know that not only are you missing aluminum but you also have a value for iron. Okay? Then the mean is not the best estimate of what the missing value is for aluminum. The best estimate is the conditional value after you take account of what you observed on iron. 03:56PM
Q When did Olsen say he unstandardized the data? 03:57PM
A Wasn't that part of the -- I have to look at the steps in my report, but as I understand the steps, he went back and unstandardized to be able to calculate those scores. Otherwise, his plots would have centered on zero.
MR. PAGE: I think we better stop here to take a tape break. 03:57PM
VIDEOGRAPHER: We are now off the Record. The time is 3:56 p.m.
(Following a short recess at 3:56 p.m., proceedings continued on the Record at 4:17 p.m.)

VIDEOGRAPHER: We are now on the Record.
The time is 4:17 p.m.
Q Dr. Cowan, was your testimony before we broke that you were able to reproduce Dr. Olsen's PCA results in its entirety? 04:18PM
A Let's just talk about SW3, if we could.
Q Yes, sir, SW3.
A Yes. 04:18PM
Q What about the run involving groundwater; were you able to reproduce that also?
A I didn't even try that.
Q So the only one you tried was SW3?
A And variations on the surface water, so SW3, SW15. We looked at some of the other things, but we only -- the only work we did was with surface water. 04:19PM
Q And when you did those, you were able to reproduce Dr. Olsen's results exactly?
A Well, within -- within rounding error like at the fourth digit. 04:19PM
Q Okay, and could you try to describe for us again where that would be in your considered materials?
A Yes, sir, and I'm sorry because I don't remember exactly how I structured the files that I provided, but the fact is that I had a number of Excel files which were outputs from either SysStat or SPSS because we ran both programs, and then you can pick up the tables that are output and stick 04:19PM
them into Excel files so that you can format them more readily.
So you will find Excel files that are labeled according to the dataset that was used, SW3, SW15 and so on, and whether or not we took logarithms or didn't take logarithms. Those are the titles of the files themselves, so they will all be Excel files, XLS, and they will have a name -- 04:20PM
Q SW3, SW15?
A Yes, sir. 04:20PM
Q Thank you. And then is it true then, sir, referring you to Exhibit No. 17, since you were able to reproduce the results for Dr. Olsen, you were able to reproduce the information that's in Exhibit 17 exactly also? 04:20PM
A Well, I don't think we tried to do all of the information that is in Exhibit 17, but we took the initial results out of that and also we got the same Eigenvalues.
Q Okay, but all of the output that you see there on Exhibit 17 would be output you could receive after you perform PCA analysis? 04:20PM
A Definitely.
Q So if you had tried to and assuming this is -- well, if you had tried to, you could have reproduced what's found in Exhibit 17 also; correct? 04:21PM
A Yes. If I had been asked to do -- to reproduce Exhibit 17, I could have done so. What I was trying to explain was that I ran the same program so they had the same outputs but I didn't necessarily save all of the outputs. I saved the parts that would be most confirmatory, like the 04:21PM

actual coefficients, so that I could compare those to the tables that were in Dr. Olsen's report.
Q And did you find that they matched? 04:21PM
A Yes, out to --
Q The fourth decimal?
A Yes, sir.
Q Dr. Cowan, would PC versus PC correlations using pairwise deletion be equal to zero? 04:22PM
A I'm sorry, I did not understand the question.
MR. PAGE: Would you restate it, please?
(Whereupon, the court reporter read back the previous question.)
A I have no idea. I heard what you said and I heard the exact same thing when she repeated it, but the question doesn't make any sense to me. 04:22PM
Q Would you turn to Pages 23 through 25 of your report?
A Sure, yes, sir. 04:22PM

0226

Q It shows some figures there, Chart 7A through 7E; is that correct?
A Yes, sir.
Q Where did the data come from that supports these figures? 04:23PM
A From SW3.
Q And this is when you -- the SW3 where you reproduced Dr. Olsen's database or Dr. Olsen's database?
A Dr. Olsen's database. 04:23PM
Q Okay. Would you read the first two sentences on page -- Paragraph 54?
A Dr. Olsen is missing a large number of observations on both calcium and alkalinity. When he is missing an observation, he substitutes the mean regardless of what he knows about the other variable. 04:23PM
Q Okay. Now, you are no longer suggesting that Dr. Olsen has substituted the mean for his missing observations, are you? 04:23PM
MR. TODD: Object to form.
A Yes, I am.
Q I thought we just had a long discussion about he didn't substitute the mean before he ran the PCA, he used pairwise deletion. 04:24PM

0227

A Well, I believe this is a matter of semantics. The pairwise deletion process after you've standardized the data is a calculation for, let's say, a pair of variables on which you have common observations. When you've got that, you are 04:24PM calculating a co-variance for the correlation to go into the correlation matrix that Dr. Olsen used. What Dr. Olsen is doing to calculate that correlation or what SysStat is doing, it calculates a co-variance in the numerator, which is the sum of 04:24PM cross products of first variable with the second variable. So let's call them Variable A and Variable B to make this simple.
So I have the sum of cross products of A with B in the numerator, and then the denominator, I have 04:25PM the sum of squares of the first variable times the sum of squares of the second variable. I take square root of that, and that ratio of the

co-variance to the square root of the product of the variances is the correlation coefficient, which is what was the input to the PCA, and since it's standardized, the means for both Variables A and Variables B are zero. So the cross product is also going to be zero. So in the numerator, I have the sum of the X, Y from the original data, plus zero. 04:25PM

In the denominator, I have the sum of the X squareds, plus zero. I have the sum Y squared, the sum of the B squareds, plus zero. So whether or not he actually put in the zeros, effectively he plugged in the zeros because it's the only way to calculate -- 04:25PM

Q I'm not talking about zeroes. We're talking about the --

MR. TODD: Would you let him finish the answer, please? Thank you. 04:26PM

Q Were you finished?

A No, but to respond to your concern, zeros are the mean. He standardized the variables. Once he standardized the variables, the mean is zero.

Q So you're not suggesting that he took the mean of the values that weren't missing and inputted those into the missing value locations? 04:26PM

A Yes, I am.

Q Oh, you're suggesting that he did -- that PCA did that, they took the mean of the data where there were not missing values and substituted that for the missing value; that's your understanding of pairwise deletion? 04:26PM

A Well, except for one thing. Keep in mind that what's actually happening is that this is happening on a temporary holding file inside of SysStat. It's working with his -- excuse me. I'm speaking to you. Okay? 04:26PM

Q You can answer the question. You talk to the judge and the jury. If I want to talk to my witness, I will. 04:27PM

A While I'm speaking?

Q You understand that? Yes, sir.

A Okay. Go ahead.

Q I can multitask here. I'll listen and talk. Is that okay with you? 04:27PM

A Sure, but this seems to be an issue that hasn't been understood, although we've covered it four times now.

Q Go right ahead with your answer. 04:27PM

A Okay. So what happens in the temporary holding file within SysStat is that it takes the data into a separate area and substitute -- it doesn't substitute. It does this calculation of the sum of X, Y divided by the sum of the X squared, sum of the Y squared, and the fact is that since Dr. Olsen had already standardized the data, it meant that all the means were zero. So essentially what was happening was that a bunch of zeros were being added in in place of the data that would have been 04:27PM

there.

Q Isn't it true that when you run SysStat with pairwise deletion, it simply ignores the missing

values and doesn't run correlations on those?
A It's mathematically the exact same thing. So if that's what Dr. Olsen did, Dr. Olsen was equivalently doing -- plugging in the means, and he needs to realize that to understand how the calculations are actually being done. 04:28PM
Q Let me ask you, where did you get the information that there is a large number of missing observations for calcium and alkalinity? 04:28PM
A Well, if you look --
Q You said that was from SW3, Dr. Olsen's database? 04:28PM
A I extracted -- yes.
Q Let me hand you what's going to be marked -- let me mark it, please, Dr. Cowan.
A Oh, I'm sorry.
Q Do you recognize Exhibit 18? 04:29PM
A No.
Q You don't?
A No.
Q Do you understand that this is the Excel database for SW3? 04:29PM

MS. COLLINS: Why don't you give him a minute to look at it.
MR. PAGE: Okay.
A Well, you know, there are a lot of databases and they all look the same. So if you'd like to allege that it is SW3, I'm willing to go on that hypothetical, but I'm just telling you that -- 04:29PM
Q Well, let's go -- you don't recognize from appearing here --
A Well, they all look the same. 04:29PM
Q Okay. Well, would you take at least a hypothetical of my representation that this is found in Dr. Olsen's materials as cross dat, underscore, water, 0427, underscore, SW, underscore, 3, dot XLS, which would be his SW3 database, correct, if that's correct? 04:30PM
A If the other parts of that are correct, then, yes, that would be his SW3 database.
Q Do you see any missing data for calcium on this database? 04:30PM
A Nope.
Q So if you were use -- if this is Dr. Olsen's database for SW3, that is Exhibit 18, you were mistaken when you said he was missing a large number of observations for calcium? 04:31PM

A I may have been, but I need to go back to my datasets to determine that.
Q Okay, and about alkalinity, would you see -- and this does -- this database does sum at the bottom. You probably already know that, but would you look for alkalinity and see how many missing observations there are for alkalinity. You can look through each one also if you'd like, sir. 04:31PM
A That's okay. It's -- if you look on the first page, you can see some blank spaces, and the summation of the number missing at the bottom is eight. 04:31PM
Q Okay. So would you -- would you consider eight to be a large number of missing data for

alkalinity? 04:31PM
A No.
Q So if you -- if this is Dr. Olsen's database for SW3, you would have been mistaken when you said there was a large number of missing observations for alkalinity; is that correct? 04:32PM
A If I had specified this, which I think I did, as SW3, then, yes.
Q Okay, and would you look at again the last page of this exhibit?
A Yes, sir. 04:32PM

0233

Q Do you see how many total missing observations are there?
A The number we discussed before, 915.
Q Would you turn now, sir, to Page 42 of your report? 04:33PM
A Yes.
Q This is where you identify the computational error that Dr. Olsen had in his computation of PC scores; is that correct?
A Yes, sir. 04:33PM
Q Is it your opinion that this computational error significantly affected Dr. Olsen's PCA evaluation?
A Yes.
Q How so? 04:33PM
A Could we turn to Page 44?
Q Uh-huh.
A Chart 9 is the plot of the two PCA scores that Dr. Olsen retains, and it has a very distinct pattern to it. 04:34PM
Q Uh-huh.
A And Dr. Olsen's claims are that values for Principal Component 1 that are greater than 1.3 are evidence of poultry litter. Okay? However, the problem is that that chart shows that there's a 04:34PM

0234

distinct correlation between Principal Component 1 and Principal Component 2, and they are not really representative of the outcome from a principal components analysis because the results are supposed to be and are required to be, by the way the 04:34PM calculation is done, uncorrelated.
So I apologize. That's my son.
Chart 10 shows what the computation would be or the outcomes would be with the calculation done correctly and within SysStat. 04:35PM
Q Do you recall how Dr. Olsen grouped by circling PC1 and PC2 in Chart 9?
A Yes --
Q Can you --
A -- vaguely. 04:35PM
Q Can you, best of your recollection, show how Dr. Olsen did that on Chart 9?
A As I recall, and I would like to point out, I'm doing this from memory from a document that I -- you know, from an analysis I did a while ago. As I 04:35PM remember, Dr. Olsen talked about this and this.
Q Okay, and would you label one PC1 and one PC2, please?
A Well, the axes already do that, so --
Q So the lower shape would be representative of 04:36PM

0235
the samples that Dr. Olsen characterized as PC1 samples?
A Well, the reason I'm struggling with this is that every point is -- reflects the position of that observation on both PC1 and PC2, so I can't -- 04:36PM
Q So those were the two groups, though, that Dr. Olsen identified?
A Those are the two groups, but I can't strictly say that one is only PC1 or PC2 because you couldn't get the groups without knowing both values. 04:36PM
Q With your recomputation in Chart 10, can you identify two groups in that plot?
A You mean the same two groups?
Q Can you identify two groups?
A No. 04:36PM
Q You can't?
A No.
Q Would you not agree with me, sir, that one group can be identified from this group here and another group can be identified with this group 04:37PM here?
A No.
MR. TODD: Just for a minute, just for the Record, the circles on Chart 10 have just been added by Mr. Page. 04:37PM

0236
MR. PAGE: I'm sure the Record will show that with the camera also, but thank you.
Q So you disagree with that characterization?
A Well, the reason that I have trouble with that characterization is that I have no idea how you 04:37PM decided to include or exclude any points from the groups or why you excluded what looks like approximately ten values to the far right or maybe thirty values to the far left.
Q So if we were going to try to identify whether 04:37PM or not those groups would represent contamination, points of contamination for poultry and wastewater treatment plant, additional analysis would have to be done?
A Well, some additional analysis -- well, again, 04:38PM I'm having trouble with your question because it's kind of broad. So the answer is, first of all, you could do an analysis to identify two groups.
Q And how would you do that?
A Without regard to -- well, there's procedures 04:38PM using, for example, cluster analysis that would consider the relative positions of all the points, and then if you specified that you wanted to have the two most homogeneous groups formed where homogeneity is defined as the distance of each point 04:38PM

0237
to the centroid of the group and heterogeneity is defined as the distance between the centroids of the two groups, there are lots of different ways to do that type of analysis without regard to what Principal Component 1 or Principal Component 2 04:39PM represent.
Q Let me ask this question then, sir: Do you know how Dr. Olsen came about identifying his two groups?
A Well, I read what he said in the report, but I 04:39PM

didn't understand how he chose a cut-off point of, for example, 1.3 to identify points to the right of that on Principal Component 1 as being related to poultry litter versus being related to anything else, and I didn't understand the choice of 1.3, especially when you look at the clustering here. I didn't understand 1.3 versus any other number.

Q And did you understand what Dr. Olsen did when he referred to a spatial analysis of these points? 04:39PM

A Well, I know what a spatial analysis is. I just described one, but there's hundreds of ways to do spatial analysis. So could you be more --

Q What about a spatial analysis by evaluating where the location of the sample was within the IRW and its potential influences by different sources, 04:40PM

0238

that type of a spatial analysis?

A But if that's the case, you could actually include that data in the analysis that's being conducted, which Dr. Olsen didn't do.

Q Did Dr. Olsen do that evaluation when he did his evaluation of PC1 versus PC2, that is, that type of a spatial analysis? 04:40PM

A Well, he says he does, but I don't have any records of the mathematics that would be required to do a rigorous spatial analysis as opposed to a by-the-eye spatial analysis. 04:40PM

Q I've just handed you what's been marked as Exhibit 19, Dr. Cowan, and I'll represent to you that we took your Chart 9 and Chart 10 and then color coded it with the type of sample that was taken so that we were able to identify the points on the plot as edge of field, Lake Tenkiller, wastewater treatment plant reference and all those. 04:41PM

A I'm confused about one thing.

Q What's that, sir? 04:42PM

A How did you take my data to create Chart 10?

Q Well, we produced your analysis and then identified those locations as the origins of the samples.

A Okay. Then if you don't mind, I'm going to 04:42PM

0239

say that that's -- I understand that you are presenting that as a hypothetical, but since I haven't been through your calculations allegedly reproducing my data, I can only guess that you did or you didn't. 04:42PM

Q Okay. Well, if you look at Chart 10, which is on Exhibit No. 19, and look at Chart 10 from your report, does it appear to you from your visual analysis that the plots on both points are the same?

A The patterns are the same, but I don't want to swear that your Chart 10 exactly reproduces my results. If you'd like to establish that as a hypothetical, I'd be happy to go from there. 04:42PM

Q That's what I'd like to do, sir.

A Yes, sir. 04:43PM

Q Okay. Would you look now at Chart No. 9 and could you identify -- this is on Exhibit 19. Would you identify and draw a circle around where the wastewater treatment plant results are located on Chart 9? 04:43PM

MR. TODD: Object to form.

A Well, I'm not sure that I can. You mean on Chart 9?
Q Yes, sir.
A Because I only see two. 04:43PM

Q Okay. Well, would you draw a circle around those two points that you've identified?
A Yes, sir. Okay. With the understanding that I'm only observing two and that there may be more that are hidden behind other points. 04:43PM
Q Okay, and can you identify the reference values on Chart No. 9 as the green triangles there; could you draw a circle around where you see the green triangles?
A Oh. Okay. I'm going to -- I will do as you asked, but I'd like to offer the same caveat that I can only identify six values. 04:44PM
Q Do you recall at this time, sir, whether there's more than two wastewater treatment plant observations on your Chart 9? 04:44PM
A Well, on my Chart 9, I don't have any. Are you asking about the Chart 9 here?
Q Well, in the dataset that you used to produce Chart 9.
A I don't recall. 04:44PM
Q Okay, and how many reference points do you find or reference samples do you find on Chart 9?
MS. COLLINS: Object to form.
Q It's the green triangles.
MS. COLLINS: Object to form. 04:45PM

A Your Chart 9 on Exhibit 19?
Q Yes, sir.
A It looks to me like it's six, but the other indicators that you've used are also triangles so I may be having trouble between dark green and dark blue. 04:45PM
Q And do you recall whether or not there were six reference samples that were for SW3 database?
A I'm sorry, again, I don't recall.
Q Okay. Would you now then draw a circle around those points that represent edge of field samples? 04:45PM
A I'm sorry, again on Chart 9?
Q Yes. Thank you.
A That's okay.
Q They're the blue diamonds I believe, sir. 04:45PM
A Yes. Well, actually I should have said this after the last grouping, too, but what you're -- let me make sure that I understand what you're asking me. You're asking me to create a group that contains -- 04:46PM
Q Edge of field.
A -- most of them or all of them.
Q All of them that you can see.
A Okay.
Q Now, when you've done that, sir, can you see 04:46PM

there's a distinction where the samples that I've represented to you are wastewater treatment plant are separated from those that are edge of field samples and separated from those that are reference samples? 04:46PM
MR. TODD: Object to form.

MS. COLLINS: Object to form.
A Well, I see that there are three non-overlapping groups that were formed by my circles. Is that your question? 04:47PM
Q Yes, sir.
A Yes.
Q And what did you understand a reference sample to be in Dr. Olsen's database?
A As I understood it, it was a sample taken where it was known or alleged that there was -- there were no poultry farms nearby. 04:47PM
Q Okay. Was it also important that there were no other sources of contamination, such as wastewater treatment plant, where the reference sample was taken? 04:47PM
A I believe that that's correct.
Q Okay, and turning now to the wastewater treatment plant samples, what was your understanding of the purpose of those samples? 04:47PM

0243

A Specifically to sample near the wastewater treatment plants to determine the composition of the surface water close to those two points.
Q Okay, sir, and then same question, was it -- is it -- was it also important that the wastewater treatment plant outfalls were sampled for wastewater treatment plant analysis; do you know that? 04:48PM
A When you mean outfalls, you mean the direct effluent from the --
Q Direct effluent from the wastewater treatment plant before it enters the receiving body of water. 04:48PM
A That's my understanding, yes.
Q Okay. So that's what the wastewater treatment plant samples were?
A Uh-huh. Well, I'm sorry. I just couldn't attest to whether it was actually at the point -- you could get the effluent as it was entering into the pipeline before it hit the water or you could get it just after it hit the water in the stream, so -- 04:48PM
Q Okay, but you believe that the wastewater treatment plant samples was one of those two?
A Yeah, and in either case it would give you very similar results. I just didn't want to give the indication that I knew exactly where the -- 04:48PM

0244

Q Fair enough. Now, when you look at the edge of field samples, do you have an understanding what the purpose of edge of field samples were?
A It was to determine what the composition of the surface water was. I'm using the term composition loosely, relative to edge of the different fields that were sampled. 04:49PM
Q Were edge of field samples in this case sampled where poultry litter was identified to have recently been applied? 04:49PM
MS. COLLINS: Object to form.
A Well, I believe that that is one set of the edge of field samples. I'm not sure that that completely describes the edge of field samples.
Q You don't recall how Dr. Olsen described that in his report? 04:49PM
A I don't recall the recency part of it.

Q You do recall that they were intended to be representative of runoff from fields where poultry litter had been applied? 04:50PM
A Yeah. The only thing I was stating was that I didn't remember whether or not it was a recent application or not.
Q Thank you. Now, let's turn to Chart 10, the second page, and I'd like for you to do the same 04:50PM

0245

thing for me, sir. Would you go about circling the wastewater treatment plant first on Chart 10 of Exhibit 17 -- excuse me, Exhibit 19.
A (Witness complied).
Q Okay, and would you circle the reference 04:50PM samples also, sir?
A Okay.
Q And it might be helpful. Would you mind labeling the wastewater treatment plant with WWTP, please? 04:51PM
A (Witness complied).
Q And maybe reference with REF.
A Yes, sir. Would you like me to do that on Chart 9 also?
Q Yeah. I think that would be helpful. Thank 04:51PM you.
A (Witness complied).
Q Then we'll call EOF maybe for edge of field.
A Sure.
Q Thank you, and then would you label edge of 04:51PM field on Chart 10 actually?
A Well, I haven't actually circled it. Would you like me to circle it first?
Q Yes, sir.
A I mean, otherwise, I'd be happy to just write 04:51PM

0246

it down but I'm not sure that would help us here. Do you want to see this?
MR. TODD: Can I see it closer?
A Certainly. Is that a blue diamond right in there? I could be wrong but -- okay. I just can't 04:52PM tell, so we'll put a dotted line over there to indicate that that might be.
Q So you've dotted a line. There may be another edge of field sample that's closer to the cluster?
A Yeah. It's just hard to tell because of the 04:53PM overlap.
Q So on Chart 10 do you see that you were able to identify three separate groups of those types of samples in your Chart 10 also?
A Yes, sir. 04:53PM
Q They're actually a little bit better or more diverse in your Chart 10 than they are in Chart 9; is that correct?
MS. COLLINS: Object to form.
MR. TODD: Object to form. 04:53PM
A I don't know about more diverse. I think what you're trying to say is that the centroids of each group are further from each other.
Q Is that true on Chart 10?
A Yes. 04:53PM

0247

Q Did you understand that Dr. Olsen intended to do a log transformation that you have identified on

Page 93 of your report and it was an inadvertent error to not do so?
A Yeah, I understood that. 04:53PM
Q Okay. Now, once you've done this transformation, is it easier or more difficult to see the group he sort of identified on Charts 9 and 10 on Exhibit 19 in your opinion?
A Well, in my opinion it's easier to see. 04:54PM
Q Once you've done the proper calculation like you did in Chart 10?
MS. COLLINS: Object to form.
A Well, that I -- that allegedly -- I didn't do the calculation. 04:54PM
Q But if that was following what you stated in your report, then it's easier to see in Chart 10; is that correct?
A That's true. I just don't want to get hung with this calculation. 04:54PM
Q If these are in fact reproductions of SW3 and I've labeled these correct, that is, my hypothetical is correct, sir, will you agree that the computational error had little effect on the ability of making these groupings? 04:54PM

0248

MS. COLLINS: Object to form.
A Well, if you are concerned only with those three groupings, it is easier to find those, but I suspect if we looked at because I just did stream high flow and HFS high flow, I would say it's harder. 04:55PM
Q Well, that's because in Chart No. 9 a lot of those are overlapping each other; is that correct?
A Yes, sir.
Q Okay. Do you have an understanding in the environment what stream high flow and HFS high flow would represent? 04:55PM
A As I understand it, those are taken in periods of time or from locations where there is a much greater flow, a faster flow than at other times. 04:56PM
Q So looking at your Chart 10, would it be fair to say that the cluster of points that are between the three groupings you've identified are likely mixtures --
MS. COLLINS: Object to form. 04:56PM
MR. TODD: Object to form.
Q -- of those types of effluents?
A If we're back to referring to the high flow?
Q Yes.
A I'm not sure what you mean by -- I'm sorry, a mixture has a very specific meaning in statistics, too, but if you're using it -- 04:56PM

0249

Q And environmental sampling --
A Yes.
Q -- where you -- yes? 04:56PM
A Yes.
MS. COLLINS: Can I ask for some clarification on the Record about this chart? I just want to make sure I understood what you -- -
MR. PAGE: Oh, what we identified it as?
MS. COLLINS: -- represented it to be, yes.
MR. PAGE: Yes. What these are are the Plots 9 and 10 from Dr. Cowan's report where we've

identified a specific origin of the sample and then provided a color and a symbol so you can show the origin of the sample. 04:57PM

MS. COLLINS: So you're saying there should be a direct correlation between Chart 9 of Exhibit 19 and Chart 9 as in Cowan's reports?

MR. PAGE: Yeah. All the dots are in the same place, intend to be in the same place and, whereas, 9 and 10 of Dr. Cowan does not distinguish between the different types of samples. What we've done is we've taken those diamonds, those blue diamonds there, and given them their specific sample of origin. 04:57PM 04:58PM

MS. COLLINS: You mean you and Dr. Olsen?

MR. PAGE: Yes.

MS. COLLINS: And that's your representation as to this? 04:58PM

MR. PAGE: Yes.

Q Can we turn to Paragraph 76 of your report, sir?

A Paragraph 76?

Q Did I say page?

A No. I just wanted to be sure I heard it correctly. I thought that's what you said.

Q That's fair enough. I mean to say Paragraph 76. I think it's on Page 33.

A I don't think I have a Page 76. 04:59PM

Q Yeah, I think it stopped at 72 we decided earlier. Page 33, Paragraph 76.

A Thank you.

Q Got to get on the same page.

MR. TODD: It's getting late.

Q What I want to focus on is Sentence No. 3. Would you read that?

A He throws away significant results that may explain patterns not found in the first two components. 04:59PM

Q What do you mean by that?

A Well, in principal components and in the results that you showed me earlier in one of the exhibits from Dr. Olsen, Exhibit 17, there are clearly more than two principal components, and the third or fourth dimension may cause the points that you just identified to be above or below the page if that was in three dimensions. Or in four dimensions, it may be on two completely different documents. So the point is that although you're looking at Principal Component 1 and Principal Component 2 in a two-dimensional space, the result that Dr. Olsen got was a five-dimensional space according to the results that were in -- that you just showed me in Exhibit 17, which comes from his report, and so I don't know whether the circle that you just had me draw is an adequate representation of grouping or whether, for example, there's three different groupings because they lie below, in the middle and above the page if I were to look at the third dimension that Dr. Olsen identified. 05:00PM 05:00PM 05:00PM

Q Okay. I want to talk about the specific words you used. We had some discussion about specific words. You state he, referring to Dr. Olsen, throws

away significant results that may explain patterns 05:01PM

0252

not found in the first two components. What do you mean by Dr. Olsen threw away significant results?

A I don't see any discussion or analysis of the other three components.

Q So it's your assumption then that he just threw them out and didn't consider them at all? 05:01PM

A Well, I didn't find reference to them in the report or in the document that you just showed me, nor in the document that we just looked at, so --

Q Is it -- I'm sorry to interrupt you. Is it your assumption that he did not consider those? 05:01PM

A Yes, because I have no indication otherwise that he included them or considered them in his discussion.

Q Okay. When you reviewed Dr. Olsen's analysis and you told him you reproduced it and could reproduce the information similar to Exhibit 17, did you go ahead and do the consideration of those results that you claim Dr. Olsen did not undertake? 05:02PM

MR. TODD: Object to form. 05:02PM

A I did some. For example, we discussed this earlier. In the tables in my report where I present all five of the principal components, and I actually do it twice, once for the USGS data and once for the non-USGS data. 05:02PM

0253

Q Okay, but when you looked at Dr. Olsen's data and the way he ran the database, did you do any evaluation of the other three principal components to see if they may be useful in explaining specific results? 05:02PM

A I'm sorry, I didn't mean to mislead you. What I was just describing was based on Dr. Olsen's data. The only thing that I did was I did the analysis three times, once for all of SW3, once for SW3 that came from the U. S. Geological Survey and once for U. S. Geological Survey -- non-U. S. Geological Survey. So I did reproduce. 05:03PM

Q Okay. So when you looked at SW3 --

A Yes, sir.

Q -- not when you separated out the USGS versus the non-USGS -- 05:03PM

A Yes.

Q -- did you evaluate the other three principal components?

A As I remember, and I think I said this somewhere in the report, the third principal component actually is driven by the commonality between the four organic constituents and the bacteria, and that isn't represented anywhere in the charts that we were discussing. 05:03PM

0254

Q Could you show me where you state that in your report?

A Sure. On Page 17, Paragraph 40.

Q Let me turn to it before you start reading to us. 05:05PM

A Sure.

Q Okay.

A As it stands, Dr. Olsen does not retain or analyze the principal component that summarizes the

bacteria. He throws it away. 05:05PM
Q Okay, and which PC is that?
A As I recall, it's No. 3.
Q Okay, and is that where you refer then on page -- is that your analysis then on Page 34 where you talk about on page -- on Paragraph 78 that using 05:05PM Dr. Olsen's methods, he would throw away the principal component that has bacteria, fecal coliform, Entero, coccus, I assume that's an abbreviation, and E. coli?
A Yes. 05:05PM
Q It's the same reference?
A I'm using the same abbreviations as I understand that were in Dr. Olsen's data, but what you just described is what I was identifying.
Q Are you claiming that Table 2 is a 05:06PM

0255

reproduction of Dr. Olsen's results, his PC results?
A As I recall, yes.
Q Well, let's turn back to Exhibit 17, if we can, and let's look at -- yeah, I guess we should do that. Instead of 17, I guess we could do this and 05:07PM then -- let's keep your Chart 2 there or Table 2. Let's keep 17, and then I'm going to Exhibit 20, and I want to see if we can determine whether or not your Table 2, which represents the throwing away of the bacteria data, actually is a reproduction of Dr. 05:07PM Olsen's results. And I was looking, for example, at Figure 6.11-10 on actually Exhibits 17 and 20.
A Uh-huh.
Q Can you see where in PC1 bacteria ranks in order and can you compare it to what you've 05:08PM represented on Table 2?
A Well, I can't compare them because Figure 16-10 -- 11-10, the one that you just referred to, specifically says at the top no rotation.
Q Okay, and this is through rotations? 05:08PM
A Yes.
Q Where do you say that this is a rotated analysis?
A Well, this section is on the distinction between rotations and not rotated. So I'm not sure 05:09PM

0256

I said it explicitly.
Q You'll agree with me, though, that if is this -- so you're saying Table 2 is rotated data; correct?
A Yes. 05:09PM
Q You'll agree with me, it doesn't match with the no rotation data; correct?
A No, it doesn't because the no rotated data loads everything onto PC1. Everything is important, whereas -- and that's the whole purpose for doing 05:10PM the rotations, to tease out what's important and what isn't.
Q So by this analysis, you're saying Dr. Olsen threw out PCs 3, 4 and 5?
A Well, you haven't -- you've handed me a 05:10PM document that has PC 3 in it from Dr. Olsen.
Q What about Exhibit No. 17?
A I don't believe it had it in it, but let me look again, please.
Q Does it not show at the bottom of Figure 6.1-1 05:10PM

that there are five principal components that are above .385 (sic) percent?
A Well, it does, but given that that's what that shows, why is it that none of the remaining charts have PC 3, 4 or 5 in it even though you just cited something that says that they're important? 05:11PM

Q Let me ask the questions, okay, Dr. Cowan?
A Sure.
Q So you'll agree with me, will you not, that under Figure 6.11-1, Dr. Olsen identified five principal components that would explain some variance other than just random? 05:11PM
A I wouldn't put it quite that way but I agree with the five.
Q Well, what does the 3.85 stand for? 05:11PM
A I'm sorry, the 3.85?
Q Yes. On Figure 6.11-1, do you know what the number 3.85 represents?
A I'm sorry. I had my thumb over it.
Q Okay. What does that represent? 05:12PM
A I believe that the 3.85 is the average amount of variance that you would explain if you accounted for each variable contributing unity to the overall variance. So in other words, if you took 1 divided by 26, I believe that's 3.85 percent. 05:12PM
Q So any principal component that fell below 3.85 would not be important to your analysis?
A Well, it may or may not be, depending on what went into it. The use of the 3.85 just means it's above or below the average amount of variance contributed by each variable. Whether or not it's important to the analysis depends on the structure of the principal component. 05:12PM
Q Well, the average would be pretty much represented by random -- random correlations; correct? 05:12PM
A Not necessarily.
Q So you're suggesting that if you have a score below 3.85 with a 26-variable PC, you would have something that would represent something other than random correlations? 05:13PM
A You could.
Q Would that be the case in this situation?
A I don't know. I didn't look at anything beyond Principal Component 5. 05:13PM
Q Why not?
A Because I only looked at -- I was critiquing the analysis that Dr. Olsen did, and it didn't seem to add a lot to the discussion to look at the other principal components out beyond No. 5, although I will say that in a couple of cases I went to 6 and 7, and you'll find that in my work papers, too. 05:13PM
Q And so are you saying that you looked through Principal Component No. 5 because that's what Dr. Olsen did, and you wanted to do the same analysis he did? 05:14PM

MR. TODD: Object to form.
A No, I didn't want to do the same analysis he did. What I was doing was critiquing the analysis he did, so I was trying to reproduce the analysis, 05:14PM

so -- but there's an easier way to answer your question.
Q Feel free.
A Thank you. Also from Dr. Olsen's outputs, from these tables, you can also compute the commonality of the variables. The commonality is a measure of how well a variable is reproduced by one, two, three, four, five, however many principal components you retain, and both Dr. Johnson and I refer to the fact that the commonality of most of the variables is reproduced pretty well with the exception of bacteria, and I think Dr. Johnson had one or two other variables that he identified. That would indicate that once you go beyond -- certainly beyond Principal Component 2 and into Principal Component 5, that you can re -- the whole purpose of principal components is to summarize the variability in the dataset. If you are able to reproduce almost all the data from the five principal components or six or seven, that means that there's nothing 05:14PM 05:14PM 05:15PM 05:15PM

0260

necessarily that is unique about the later principal components. However, in other situations and analyses that I've done, you can wind up with one variable that only explains itself, and so it will -- you don't get an explanation of the commonality or improvement in the commonality until you get out to, let's say, Principal Component 10 because that variable doesn't kick in until then.
Q So in this particular case, how many principal components would you think were important to take a look at to your analysis?
A I believe that five.
Q Five?
A Was the number that we all agreed to.
Q Okay, and it's your testimony that Dr. Olsen threw out Principal Components 3, 4 and 5 when he did his analysis?
A Well, if you'd like to put it another way, I don't recall seeing a discussion of Principal Components 3, 4 and 5 as to their importance or how they would change the structural composition of the charts that we discussed earlier, Charts 9 and 10.
Q Let me hand you what's been marked as Exhibit 21, and I'll ask if you can identify that for me, please, sir. 05:15PM 05:15PM 05:16PM 05:16PM

0261

A This looks like a series of plots of the cross plots between PC1, 2, 3, 4 and 5, which would give us ten plots. The first few are unrotated, and the second set seem to be the Varimax -- well, there's some Varimax rotation and some Equamax, some Portamax.
Q Does this show an analysis of PCs 1 through 5?
A No. I'm going to say the same thing that I said early in the day. This is a set of plots. It doesn't do anything other than present what the outcomes are, but that's not an analysis.
Q Was this material in Dr. Olsen's considered materials?
A I believe it was.
Q So you're suggesting that this type of analysis isn't a consideration of PCs 3, 4 and 5? 05:16PM 05:17PM 05:17PM

A Well, I'm saying it's not an analysis. An analysis is where you consider what it is you're looking at and draw some conclusions. There's no conclusions here. There's just a set of charts. 05:18PM

Q Okay. So you're concerned that he didn't write down what he saw when he looked at the charts?

MR. TODD: Object to form.

Q Is that your criticism?

A Well, no. It goes beyond that. He did the 05:18PM

0262

charts, but I don't know anything about what he concluded or whether or not there's a conclusion to be drawn from this. This is just a -- I mean, he could have included a picture from, you know, Manet's Givenchy, and that would have been every bit as telling as these. 05:18PM

Q So it's your testimony that this doesn't help -- these types of plots do not help you evaluate the importance of PCs 3, 4 and 5?

A No. That's the exact opposite of what I just said. I said that they do help with the analysis. The problem is that I have no idea what Dr. Olsen did with these. As far as I know, he didn't look at them, and they were put in by a research assistance. They're not referenced in the report. 05:18PM 05:19PM

Q They were in his considered materials, so doesn't that by definition mean that Dr. Olsen considered these when he did his analysis for his report?

MR. TODD: Object to form. 05:19PM

A Well, it seems to me that you're elevating the term considered to be something beyond the fact that he glanced at them.

Q You don't know what he did with these reports?

A I have no idea because he didn't document what 05:19PM

0263

he did.

Q And let me ask you one more question before we break for the evening. On Table 2, what rotation was used in Table 2?

A Varimax. 05:19PM

MR. PAGE: Let's break for the evening.

VIDEOGRAPHER: We are now off the Record. The time is 5:19 p.m.

(Whereupon, the deposition was recessed at 5:19 p.m.) 05:19PM

0264

SIGNATURE PAGE

I, Charles Cowan, PhD, do hereby certify that the foregoing deposition was presented to me by Lisa A. Steinmeyer as a true and correct transcript of the proceedings in the above styled and numbered cause, and I now sign the same as true and correct.

WITNESS my hand this _____ day of _____, 2009.

_____
CHARLES COWAN, PhD

SUBSCRIBED AND SWORN TO before me this _____ day of _____, 2009.

_____
Notary Public

My Commission Expires:
_____

C E R T I F I C A T E

STATE OF OKLAHOMA)
) ss.
COUNTY OF TULSA)

I, Lisa A. Steinmeyer, Certified Shorthand Reporter within and for Tulsa County, State of Oklahoma, do hereby certify that the above named witness was by me first duly sworn to testify the truth, the whole truth and nothing but the truth in the case aforesaid, and that I reported in stenograph his deposition; that my stenograph notes were thereafter transcribed and reduced to typewritten form under my supervision, as the same appears herein.

I further certify that the foregoing 264 pages contain a full, true and correct transcript of the deposition taken at such time and place.

I further certify that I am not attorney for or relative to either of said parties, or otherwise interested in the event of said action.

WITNESS MY HAND AND SEAL this 26th day of February, 2009.

_____
LISA A. STEINMEYER, CRR
CSR No. 386

CORRECTIONS TO THE DEPOSITION OF
CHARLES COWAN, PhD
Volume I

PAGE AND LINE NUMBER	CORRECTION